Exploring the repeat protein universe through computational protein design

Authors
TJ Brunette* 1,2, Fabio Parmeggiani* 1,2, Po-Ssu Huang* 1,2, Gira Bhabha 3, Damian Ekiert 4, Susan E, Tsutakawa 5, Greg L. Hura 6, John A. Tainer 5,6 and David Baker 1,2,7

*These authors contributed equally to this work

Affiliation
1 Department of Biochemistry, University of Washington, Seattle, WA 98195, USA
2 Institute for Protein Design, University of Washington, Seattle, WA 98195, USA
3 Department of Cellular and Molecular Pharmacology, UCSF, San Francisco, CA 94158, USA
4 Department of Microbiology and Immunology, UCSF, San Francisco, CA 94158, USA
5 Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
6 Department of Molecular and Cellular Oncology, The University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Blvd., Houston, Texas 77030, USA
7 Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195, USA

Contents

Supplementary Discussion 1 | Computational protocol

We have developed a method for construction of Designed Helical Repeats (DHRs) depicted in Extended Data Fig. 1 and described below. We designed proteins based on repeating units formed by two helices and two loops. For all proteins this design process was completely automated and no manual refinement was involved. Using this protocol 69 proteins with diverse architectures were selected from the *in silico* candidates. For 14 models, an additional version that included disulphide bonds was selected, for a final list of 83 proteins that were experimentally tested. This design method has progressed over the duration of this research and only the final design method is described below. The status of the computational design method when each design was experimentally tested is illustrated in Supplementary Table 1.

Each of the following sections describes one step in Rosetta_examples and corresponds to the flow chart in Extended Data Fig. 1. The fraction of models passing each design stage and their distribution according to helix and loop length are shown in Extended Data Fig. 2. The Rosetta design code for each step is provided as Supplementary Material: Rosetta_examples, which requires Rosetta git version c876538.

1 Backbone Design

The backbone design stage employs a simplified side chain representation (centroid)[1]. The backbone assembly procedure begins by picking fragments harvested directly from a non-redundant set of structures from PDB[2]. The fragments contain only residues that fall into the space of phi-psi backbone angles of either helices or loops depending on the desired secondary structure. Loop fragments could be further specified to fall within desired ABEGO bins[3] as described by Koga *et al.* [4].

The fragments were assembled using a Monte-Carlo sampling procedure that was initialized with ideal-helices and extended loops. After every fragment sampling step, which was allowed only in the first repeat unit and at the junction between the first and the second units, the change was propagated to all downstream repeats and scored. The score function we used considered van der Waals interactions, packing, values of backbone dihedral angles, and radius of gyration (RG) that was applied to only the first and second repeat-unit (RG-local). The RG term promotes the formation of globular proteins so applying RG to the whole model produced only highly curved structures. The sampling procedure in the database used 1500 Monte Carlo fragment insertions and was further improved to 3200 steps ordered as following: 100 Monte Carlo moves with 9 residue fragments then 100 moves with 3 residue fragments, both allowed only in loops. The loop sampling was followed by 1500 moves with 9 residue fragments and 1500 moves with

3 residue fragments, both in helices and loops (improved sampling). The improvements resulted in a 3.3 times increase of acceptance at the centroid stage. The backbone was represented as poly-tyrosine during the centroid building, maintaining enough space within the core to accommodate both small and large side chains in the design step.

Using this procedure we designed 2.88 million backbones by making 500 structures for each of 5776 different secondary structure combination.

***2 Backbone quality filter*: RMSD loop threshold and motif score**

Designed backbones were screened for native-like features. First, loops were checked so that there was at least one 9-residue fragment from the PDB database within 0.4Å RMSD on every position in the structure (RMSD loop threshold). To do this we used the worst9mer filter in Rosetta [1]. Second, the design-ability of each residue was measured by the number of pairwise side chain interactions observed in the PDB database, considering the backbone position of the two residues involved (motif score, unpublished results). Backbones with fewer then 1.5 interactions per residue were filtered out. Of the 2.88 million initial backbones 66,776 structures passed these filters.

3 Sequence design - Fast

Starting from the filtered backbone conformations, we used one pass of Rosetta design[5] to generate repeated sequences.

4 Packing filters – Low threshold

After completing sequence design the models were filtered out if the helices were either too far apart, creating cavities in the core (poor Rosetta holes[6] score, > 1.75), or too close together with an alanine-rich unspecific core packing (%alanine residues > 25%). Of the 66,7776 structures that passed centroid 11,243 pass this filter.

5 Structure Profile

The structure profile biases the sequence composition towards the sequences in native proteins with similar local structure. To construct the structural profile, the sequences from the closest 100 9-residue fragments within 0.5Å RMSD to the designed structure were used. The code to construct the structural profile is included with Rosetta as generate_struct_profile.rb in tools/fragment_tools/pdb2vall. The structure profile was used in the same way as the sequence profile described by Parmeggiani *et al.*[7]

6 Sequence design – multipass

Starting from the filtered backbone conformations, we used Rosetta design to generate repeated sequences while minimizing the overall energy [4,8], increasing core packing as measured by Rosetta holes [6] and improving the psipred secondary structure prediction [9]. After the first round of sequence refinement the N and C terminal repeats (capping repeats) display exposed hydrophobic residues. The sequence design procedure was-rerun for these repeats without a symmetric sequence to introduce polar amino acids.

7 Packing filters –High threshold

After completing sequence design the models were filtered out for poor packing. (holes score, < 0.5). After this stage we obtained 1980 structures.

8 Exploration of the energy landscape

The designs were validated using Rosetta *ab initio* structure prediction using Rosetta@Home [10,11]. In Rosetta *ab initio* prediction the energy landscape is explored using independent simulations starting from an extended structure. The distribution of the simulation results is expressed in terms of energy and distance from the target fold as root mean square deviation (RMSD). A successful design produces a distribution in the shape of a funnel with the minimum corresponding to low energy and low RMSD models and no alternative minima.

For each structure, seven family members were made from the same topology, some with increased hydrogen bond potential. Proteins where multiple family members had successful simulations were selected. The member of the family with the tightest folding funnel was chosen by visual inspection and the corresponding gene was ordered for experimental testing. Extended data Fig. 3 illustrates the folding funnel and sequence diversity for one topology.

For the database we have 761 structures that have at least one family member < 3.0 RMSD from the design.

9 Add disulphides

Additional, versions with stabilizing inter-repeat disulphide bonds were also generated. Potential disulphides were scored using RosettaRemodel[12] and if the disulphide score was < 0 they were considered.

Time estimates

Backbone design: on a single core of a Xeon E5-2650 took 104.5 seconds to build a structure with a 19H-2L-20H-3L topology, the median topology in the database. With an average design time of 104.5 seconds per model, it would take 3493 compute days on a single core to generate the 2.8 million structures.

Sequence design – multipass: the multipass design of sequence and capping residues takes 2.1 hours for a model with 17 length helices and 3 length loops on a single core of a Xeon E5-2650.

Exploration of the energy landscape: on a single core of a Xeon E7-2850 @ 2.00 GHZ a model with 17 residues helices and 3 residues loops is produced in 19.7 minutes. Where the computation was run on Rosetta@Home, the average was 26.7 minutes. With 7 sequences per family and a minimum of 1000 models to suitably explore the landscape it would take 130 compute days per structure.

Supplementary Table 1 | Detailed protocol used for each design

	asymmetric topology (a)	structure profile	RMSD loop threshold	disulphide threshold < 0.0 (b)	energy landscape exploration	no csts (c)	motif score	ABEGO type (d)	improved sampling (d)
DHR 1-4,9									
DHR 5-8,10									
DHR 11-18									
DHR 19-30									
DHR 56-83									
DHR 31-55									

Dark grey cells indicate the computational method variants employed for groups of designs. The database described in section 1 of the supplementary corresponds to the technique used to make DHR56-83. (a) For DHR1-4,9,11-18 the repeat backbone at the centroid level was symmetric, with first and second helices and first and second loops having the same length and conformation. The design stage was not restricted, introducing structural and sequence variability between the two halves of the repeat. (b) A higher disulphide score threshold of 1.5 was initially used which resulted in many disulphide-containing structures being non-functional. (c) We initially used ambiguous constraints between the helices. Ambiguous constraints gave a score bonus to centroid models when a helix was within 10 Å to a helix in adjacent repeat. These constraints were found to disrupt loops and result in many structures that would not fold during simulations. (d) DHR31-55 contained a displacement between helices, which resulted in highly twisted structures. This displacement was observed when the ABEGO loop types GBB and BAB were coupled with specific helix lengths. An improved sampling strategy with increased number of Monte Carlo steps was also used in these cases. The other method variants indicated in the columns are described in supplementary section 1 (names underlined).

Supplementary Discussion 2 | Geometrical parameters of Designed Helical Repeat proteins

1) Global parameters
2) Extracting parameters from naturally occurring repeats
3) Local parameters

1) Global parameters

Class 3 repeat proteins, as described by Kajava A.[13], form solenoid structures that can be described in term of global helical parameters that relate the position of one repeat to the next one: radius (*r*), twist or angle between adjacent repeats around the helical axis (twist, ω) and translation between adjacent repeats along the helical axis (*z*).

Parameters for Designed Helical Repeat proteins (DHRs) and crystal structures are indicated in Supplementary Table 2, together with the Cα RMSD values for the complete proteins. The parameters were measured on the two central repeats using the RepeatParameter filter available in Rosetta.

Radius and twist are inversely correlated and their distribution of whole set describes a hyperbolic shape, which can be represented as two symmetric ones, when considering the handedness of the superhelix in the ω value as in Fig. 2b (+ right handed, - left handed). Handedness refers to the superhelix described by the center of mass of the repeats. *z* is broadly distributed, with maximum values around 16 Å (Extended Data Fig. 4).

2) Extracting parameters from naturally occurring repeats

A set of alpha-helical solenoid proteins were curated from the repeatsDB (category III.3.) [14] to remove both proteins that had above 90% sequence identity [15,16] and previously designed repeat proteins. After curation, 258 proteins remained out of 923. We then automatically extracted repeat units, which consisted of 3 subsequent repeats, that differed by less than 3 residues in length and had a high degree of structural similarity as measured by having a TM score [17] of greater than 0.75. The requirement of high structural similarity cut down the number of repeat proteins to 81. Repeat units were identified by the method described by RAPHAEL [18] implemented in Rosetta and improved. This method measures the distance from residues in the protein to random points placed around the protein. Equally spaced inflection points, where a residue was furthest or closest to these random points indicated the start of a repeat.

We found that inflection points occurred at random in repeat protein loops. To ensure each repeat was cut at the same location, the first residue in each repeat was chosen to be the loop-helix transition closest to the transition point. The code for this is available as

extractNativeRepeats in Rosetta after git branch c876538. After locating repeats we assigned the class name of each repeat based on the PDB assignment in the Pfam database[19]. The Rise/Omega/Twist parameters were calculated by superimposing the first repeat-unit onto the second using TM-align[17] then calling the parameter calculators and averaging the values within the same protein. This approach does not provide an extensive coverage of all the possible curvatures for each family but an indication of the protein average values.

3) Local parameters

Local parameters describe the helix-helix interactions and, due to the repeating structures, only two interactions are needed to capture the local geometry: helix1.1-helix1.2 within a repeat and helix1.1-helix2.1 between first and second repeat. Angle between helices and distance between helix centers of mass were used as parameters, extracted with a modified version of the publicly available script http://www.pymolwiki.org/index.php/AngleBetweenHelices. Secondary structure definition were assigned using DSSP[20]. Values are reported in Supplementary Table 3 for designs and crystal structures. For the two central repeats, all atoms RMSDs between crystal structures and design are reported. Repeat handedness, as defined by Kobe and Kajava[21], indicates the rotation of the main chain going from the N- to the C-terminal around the axis connecting the repeat centers of mass.

Supplementary Table 2 | Global geometric parameters.

	Secondary structure (residues)						Global parameters design				Global parameters crystal				
DHR	Repeat length	helix1	loop1	helix2	loop2	disulphide	rise (z) (Å)	radius (r) (Å)	twist (w) (rad)	hand	rise (z) (Å)	radius (r) (Å)	twist (w) (rad)	hand	Backbone RMSD (Å)
1	44	18	4	18	4	yes	8.17	3.29	0.52	L					
2	44	18	4	18	4		8.57	0.52	0.43	R					
3	42	19	2	19	2	yes	9.00	0.78	0.57	L					
4	50	21	4	21	4	yes (9 with S-S)	8.58	4.93	0.58	L	8.66	2.43	0.54	L	1.50
5	50	21	4	21	4	yes (10 with S-S)	4.99	163.77	0.06	R	6.14	80.88	0.11	R	1.73
6	42	17	4	17	4	yes	6.62	31.58	0.25	R					
7	42	18	3	18	3	yes	6.99	10.51	0.66	L	7.35	10.57	0.67	L	1.30
8	45	19	3	19	4		8.97	2.19	0.40	R	8.53	3.44	0.39	R	2.28
9	50	21	4	21	4		8.33	4.59	0.58	L					
10	50	21	4	21	4		9.22	138.20	0.03	R	5.46	79.97	0.12	R	1.79
11	44	18	4	18	4	yes (2 with S-S)	8.48	0.55	0.42	R					
12	36	15	3	15	3	yes	8.98	0.60	1.03	L					
13	36	15	3	15	3		9.35	1.89	1.02	L					
14	40	17	3	17	3		9.34	23.62	0.17	L	3.00	655.10	0.02	L	2.38
15	46	20	3	20	3	yes	9.11	0.58	0.43	R					
16	44	20	2	20	2	yes	8.62	0.71	0.50	R					
17	54	25	2	25	2	yes	8.95	44.91	0.10	R					
18	60	26	4	26	4	yes	8.82	1.08	0.36	R	8.79	4.16	0.36	R	1.21
19	38	16	3	16	3		9.60	9.83	0.23	R					
20	38	16	3	16	3		9.08	4.23	0.48	R					
21	38	16	4	15	3		9.35	7.26	0.41	R					
22	40	17	3	17	3		8.97	3.96	0.37	R					
23	40	17	3	17	3		8.97	15.68	0.30	R					
24	46	20	3	20	3	yes	9.44	5.67	0.31	R					
25	45	20	3	19	3		9.19	4.36	0.37	R					
26	52	23	3	24	2	yes	9.26	1.89	0.32	R					
27	54	24	3	24	3		8.09	10.02	0.43	R					
28	55	24	2	25	4		9.57	10.26	0.20	L					
29	53	24	3	23	3		8.99	8.72	0.30	R					
30	55	24	3	25	3		9.29	4.93	0.37	R					
31	41	17	3	18	3		9.90	8.71	1.15	R					
32	41	17	3	18	3		10.20	8.89	1.05	R					
33	41	17	3	18	3		11.32	9.20	0.89	R					
34	41	17	3	18	3	yes (33 with S-S)	11.30	8.81	0.88	R					
35	41	17	3	18	3		12.21	1.78	0.50	R					
36	41	17	3	18	3		12.29	7.28	0.94	R					
37	41	17	3	18	3		9.51	10.48	0.87	R					
38	41	17	3	18	3	yes (37 with S-S)	9.59	10.41	0.86	R					
39	42	17	3	19	3		11.68	9.24	0.62	R					
40	43	17	3	20	3		11.41	2.77	0.55	R					
41	57	24	3	27	3	yes	12.11	5.17	0.36	R					
42	57	24	3	27	3		7.58	27.02	0.33	R					
43	57	24	3	27	3		8.16	17.72	0.40	R					
44	35	16	3	14	2		9.45	23.98	0.33	R					
45	35	16	3	14	2		5.81	24.08	0.35	R					
46	39	16	3	17	3		7.65	22.79	0.37	R					
47	39	16	3	17	3	yes (46 with S-S)	8.05	25.79	0.31	R					
48	40	17	3	17	3		7.54	26.84	0.24	R					
49	40	18	2	17	3		2.31	54.74	0.18	L	3.88	102.53	0.09	L	0.87
50	43	17	4	19	3		0.49	45.48	0.27	R					
51	43	17	4	19	3		7.44	26.16	0.28	L					
52	43	17	4	19	3	yes (51 with S-S)	8.58	22.87	0.20	L					
53	43	19	3	17	4		9.92	24.75	0.29	R	6.87	29.40	0.37	R	1.33
54	43	20	3	18	2		8.76	15.10	0.26	L	8.67	16.08	0.25	L	0.93
55	46	20	3	20	3		3.30	53.00	0.18	R					
56	46	20	3	20	3	yes (55 with S-S)	3.47	51.28	0.19	R					
57	47	20	3	21	3		9.20	18.46	0.26	R					
58	47	20	3	21	3	yes (57 with S-S)	8.84	22.57	0.23	R					
59	48	20	3	22	3		8.05	9.56	0.38	R					
60	33	16	2	12	3		5.26	18.45	0.53	R					
61	33	16	2	12	3	yes (60 with S-S)	4.77	19.96	0.50	R					
62	43	20	3	17	3		4.68	27.38	0.37	R					
63	54	24	3	25	2		5.66	27.68	0.32	R					
64	57	24	3	27	3		4.52	27.91	0.37	R	3.04	32.27	0.36	R	1.54
65	57	24	3	27	3	yes (64 with S-S)	3.73	26.63	0.41	R					
66	54	26	3	23	2		6.83	13.24	0.49	R					
67	60	27	3	27	3		9.20	50.96	0.08	R					
68	61	27	3	27	4		3.22	24.75	0.46	R					
69	62	27	3	28	4		5.47	17.89	0.58	R					
70	50	19	4	25	2		1.99	28.94	0.41	R					
71	50	20	3	25	2		4.82	43.07	0.23	R	5.12	49.02	0.21	R	0.67
72	37	20	4	11	2		1.33	27.28	0.44	L					
73	50	21	2	25	2		0.42	12.18	0.95	R					
74	55	22	4	27	2		4.84	19.87	0.46	R					
75	51	26	3	20	2		2.23	26.87	0.47	L					
76	55	26	3	23	3		7.31	30.53	0.27	R	5.92	25.09	0.37	L	1.73
77	51	27	2	19	3		2.30	20.24	0.68	L					
78	51	27	2	19	3	yes (77 with S-S)	2.17	19.77	0.70	L					
79	56	27	2	23	4		6.86	22.51	0.42	R	8.61	23.89	0.32	R	1.04
80	57	28	2	24	3		0.84	12.10	1.01	R					
81	57	28	2	24	3	yes (80 with S-S)	0.60	12.27	1.00	R	0.87	12.29	1.00	R	0.65
82	58	28	2	24	4		7.41	18.85	0.47	R					
83	58	28	2	24	4	yes (82 with S-S)	7.61	18.88	0.44	R					

For each design, repeat size, secondary structure length and presence of disulphide bonds are indicated. Designs and crystal structures are compared based on global parameters *r*, *z*, *ω* and handedness of the super helix. Parameters are calculated within the two central repeats. Backbone RMSD refers to the whole protein.

Supplementary Table 3 | Local geometric parameters

	Secondary structure (residues)					Local parameters design					Local parameters crystal					
DHR	length	helix1	loop1	helix2	loop2	angle H1.1-H1.2 (deg)	dist H1.1-H1.2 (Å)	angle H1.1-H2.1 (deg)	dist H1.1-H2.1 (Å)	hand	angle H1.1-H1.2 (deg)	dist H1.1-H1.2 (Å)	angle H1.1-H2.1 (deg)	dist H1.1-H2.1 (Å)	hand	RMSD (Å)
1	44	18	4	18	4	156.8	10.75	29.67	8.48	R						
2	44	18	4	18	4	174.28	11.11	24.44	8.83	L						
3	42	19	2	19	2	161.77	9.98	32.68	9.35	R						
4	50	21	4	21	4	157.54	11.06	31.44	8.62	R	162.43	11.85	31.08	8.97	R	0.78
5	50	21	4	21	4	134.73	10.9	2.29	10.18	L	144.04	11.63	6.51	10.19	L	1.66
6	42	17	4	17	4	130.79	11.12	12.64	9.52	L						
7	42	18	3	18	3	174.62	8.69	37.03	8.06	R	175.48	8.8	38.16	8.4	R	0.39
8	45	19	3	19	4	168.52	9.89	21.33	9.51	L	166.97	10.5	22.09	9.03	L	1.51
9	50	21	4	21	4	158.68	11.66	30.35	8.02	R						
10	50	21	4	21	4	137.21	11.35	1.85	10.36	L	143.83	11.53	6.38	10.2	L	1.57
11	44	18	4	18	4	174.52	11.12	24.3	8.82	L						
12	36	15	3	15	3	141.47	9.3	57.87	10.02	R						
13	36	15	3	15	3	139.33	10.33	55.94	10.41	R						
14	40	17	3	17	3	169.4	10.87	9	10	R	155.64	11.14	0.35	11.28	R	1.64
15	46	20	3	20	3	168.37	11.6	24.49	9.41	L						
16	44	20	2	20	2	163.78	10.66	28.58	9.03	L						
17	54	25	2	25	2	176.63	9.4	5.31	9.92	R						
18	60	26	4	26	4	164.82	8.95	20.21	8.94	L	166.67	9.25	19.99	8.78	L	0.61
19	38	16	3	16	3	170.03	9.38	13.54	10.15	L						
20	38	16	3	16	3	169.89	9.32	27.33	10	L						
21	38	16	4	15	3	167.52	9.16	21.84	9.41	L						
22	40	17	3	17	3	155.85	12.04	20.61	9.01	L						
23	40	17	3	17	3	160.85	10.99	14.51	9.65	L						
24	46	20	3	20	3	175.06	9.25	16.65	9.48	L						
25	45	20	3	19	3	171.37	10.52	21.42	9.63	L						
26	52	23	3	24	2	176.91	9.94	18.68	9.46	L						
27	54	24	3	24	3	166.13	10.63	21.81	10.09	L						
28	55	24	2	25	4	172.03	10.6	11.42	9.67	R						
29	53	24	3	23	3	177.57	10.36	16.18	9.13	L						
30	55	24	3	25	3	159.21	12.09	19.87	9.38	L						
31	41	17	3	18	3	133.32	8.73	31.18	11.06	L						
32	41	17	3	18	3	134.3	9.08	30.84	11.01	L						
33	41	17	3	18	3	138.39	9.15	31.37	12.29	L						
34	41	17	3	18	3	137.61	9.07	29.33	11.99	L						
35	41	17	3	18	3	136.85	10.75	25.65	12.37	L						
36	41	17	3	18	3	140.84	10.35	41.58	13.69	L						
37	41	17	3	18	3	143.68	8.91	21.15	10.8	L						
38	41	17	3	18	3	143.75	8.94	21.33	10.82	L						
39	42	17	3	19	3	138.82	9.73	31.66	13.03	L						
40	43	17	3	20	3	152.32	11.83	31.58	12.44	L						
41	57	24	3	27	3	166.91	9.43	18.47	12.35	L						
42	57	24	3	27	3	167.37	10	16.6	12.83	L						
43	57	24	3	27	3	169.58	11.15	22.58	11.65	L						
44	35	16	3	14	2	149.56	9.71	17.62	12.78	R						
45	35	16	3	14	2	160.61	10.12	19.54	9.85	L						
46	39	16	3	17	3	151.16	9.97	20.31	12.23	R						
47	39	16	3	17	3	153.86	10.02	17.2	12.19	R						
48	40	17	3	17	3	148.53	11.62	13.56	10.09	L						
49	40	18	2	17	3	171.04	9.84	4.95	10.68	R	172.61	10.14	1.19	10.14	R	0.43
50	43	17	4	19	3	160.01	8.66	7.63	11.06	R						
51	43	17	4	19	3	160.09	9.61	9.43	11.37	R						
52	43	17	4	19	3	160.81	9.44	8.84	10.08	R						
53	43	19	3	17	4	154.71	10.08	16.36	12.8	R	147.57	9.78	24.51	13.23	R	0.47
54	43	20	3	18	2	170.64	9.19	14.7	9.28	R	166.61	9.5	15.71	9.6	R	0.54
55	46	20	3	20	3	169.91	9.83	7.17	10.91	L						
56	46	20	3	20	3	170.09	9.84	7.56	10.91	L						
57	47	20	3	21	3	167.95	10.21	14.85	10.49	R						
58	47	20	3	21	3	168.06	10.27	13.27	10.47	R						
59	48	20	3	22	3	177.91	9.69	19.14	9.73	L						
60	33	16	2	12	3	169.11	8.87	21.61	9.54	L						
61	33	16	2	12	3	172.51	8.71	18.7	9.54	L						
62	43	20	3	17	3	168.49	7.63	15.78	12.08	R						
63	54	24	3	25	2	164.1	8.58	12.75	11.73	R						
64	57	24	3	27	3	167.98	9.42	13.33	10.37	R	167.72	10.06	11.33	10.97	R	0.83
65	57	24	3	27	3	167.31	9.53	12.65	10.36	R						
66	54	26	3	23	2	164.48	9.68	23.75	8.13	L						
67	60	27	3	27	3	178.59	9.72	4.07	10.19	L						
68	61	27	3	27	4	166.7	8.97	11.61	10.38	R						
69	62	27	3	28	4	163.82	8.34	21.36	10.07	R						
70	50	19	4	25	2	165.2	9.56	10.82	10.71	R						
71	50	20	3	25	2	163.25	9.08	7.76	10.3	R	167.15	9.28	7.31	10.74	R	0.29
72	37	20	4	11	2	174.75	8.21	15.03	10.75	L						
73	50	21	2	25	2	162.52	9.25	17.32	8.51	R						
74	55	22	4	27	2	168.31	9.21	20.15	12.06	L						
75	51	26	3	20	2	170	8.06	14.85	11.41	L						
76	55	26	3	23	3	169.93	9.68	14.46	10.36	L	164.79	10.14	18.85	9.91	L	1.23
77	51	27	2	19	3	165.14	9.13	24.04	11.49	L						
78	51	27	2	19	3	165.31	9.07	24.48	11.33	L						
79	56	27	2	23	4	175.56	9.45	22.63	10.36	L	172.81	9.59	17.92	10.56	L	0.98
80	57	28	2	24	3	161.22	7.9	28.78	14.76	R						
81	57	28	2	24	3	161.44	7.89	28.66	14.75	R	161.91	8.04	27.47	14.81	R	0.42
82	58	28	2	24	4	163.71	7.95	24.37	12.57	R						
83	58	28	2	24	4	163.37	8.05	23.8	12.26	R						

For each design, repeat size and secondary structure length is indicated. Local parameters (angle and distance between helix1 of repeat1 and helix2 of repeat 1; angle and distance between helix1 of repeat1 and helix1 of repeat 2) and handedness of the repeat for designs and crystal structures are indicated. Angles, distances and RMSD are calculated within the two central repeats.

Supplementary Discussion 3 | Structure and sequence comparison

Structural comparison of experimentally validated designs with representative repeat proteins from repeatDB[14] revealed that DHRs cluster in different families than the existing repeat proteins (Extended Data Fig. 8). Additionally, designs are equally distributed between right-handed and left-handed architecture, as referred to the repeat handedness (see local parameters above), in contrast to known alpha helical repeat proteins, which are mostly right-handed. This result indicates that the handedness observed is not an intrinsic limitation of repeat proteins structures but the result of a bias during evolution.

At the sequence level, DHRs are expected to be similar to repeat proteins with high units formed by two helices, like tetratricopeptide repeats (TPR). The results of BLAST[22] searches over non-redundant NR NCBI database with an e-value cutoff of 0.0001 indicate that no characterized repeat protein was identified among the top hits (Supplementary Table 4).

HHsearch[23] comparison of single repeat sequences with the whole Pfam profile database[19], shows a similar trend (Supplementary Table 4), revealing that the hits are generally only very weak (Prob, E-value, P-value, score) and often rely only on a subset of position (Cols compared to repeat size). For several designs the first repeat protein hit is not among the first hits (see rank) and in some cases no similarity to repeat protein Hidden Markov Models is detected (n.d.), As reference helical proteins for this search, the ANK and TPR clans were considered, as well as spectrin and mterf families. Most helical repeat proteins (e.g. ARM, PPR, HAT, PUF) are included in the TPR clan.

Supplementary Table 4 | Sequence comparison

Blast				HHsearch								
DHR	top hits	Score (bits)	E-value	rank	PFAM	family	Prob.	E-value	P-value	score	cols	size
1	n.d.			2	PF13431	TPR_17	61.9	1.1	7.60E-05	19.3	20	44
2	n.d.			8	PF07719	TPR_2	18.7	19	0.0013	13.4	13	44
3	n.d.			1	PF02854	MIF4G	55.4	1.7	0.00012	21.7	28	42
4	n.d.			41	PF02536	mTERF	44.2	3.4	0.00023	24.2	26	50
5	n.d.			33	PF12862	Apc5	26.2	11	0.00071	18.9	18	50
6	n.d.			1	PF14938	SNAP	86.2	0.073	4.90E-06	30.9	37	42
7	n.d.			n.d.								42
8	n.d.			n.d.								45
9	n.d.			49	PF02536	mTERF	40.8	4.2	0.00028	23.8	26	50
10	n.d.			n.d.								50
11	n.d.			7	PF07719	TPR_2	19.7	17	0.0012	13.5	13	44
12	ref|XP_003971230.1| PREDICTED: procollagen C-endopeptidase enhan...	57	3.00E-06	n.d.								36
13	gb|ETN86505.1| hypothetical protein NECAME_16287 [Necator americ...	54	3.00E-05	10	PF13812	PPR_3	43.2	3.7	0.00025	15.2	26	36
14	n.d.			7	PF12862	Apc5	33.8	6.4	0.00043	19	30	40
15	n.d.			5	PF05918	API5	27.4	9.8	0.00066	26.3	25	46
16	n.d.			7	PF12862	Apc5	32.2	7.1	0.00048	19.1	23	44
17	ref|XP_004079406.1| PREDICTED: laminin subunit alpha-3-like [Ory...	56	2.00E-05	n.d.								54
18	ref|XP_002165606.2| PREDICTED: uncharacterized protein LOC100202...	62	2.00E-07	1	PF13174	TPR_6	51	2.3	0.00015	17.1	12	60
19	ref|XP_001634212.1| predicted protein [Nematostella vectensis] >...	60	6.00E-07	2	PF12862	Apc5	73.3	0.45	3.00E-05	23.3	26	38
20	n.d.			6	PF12862	Apc5	33.2	6.7	0.00045	18.7	25	38
21	n.d.			6	PF13525	YfiO	31.1	7.6	0.00051	20.6	23	38
22	n.d.			n.d.								40
23	ref|XP_005533922.1| PREDICTED: LOW QUALITY PROTEIN: staphylococc...	54	3.00E-05	n.d.								40
24	ref|WP_003781896.1| anaphase-promoting protein subunit 3 [Actino...	52	2.00E-04	12	PF00637	Clathrin	37.7	5.1	0.00034	20	24	46
25	n.d.			8	PF11865	DUF3385	31.5	7.4	0.0005	21.5	28	45
26	ref|WP_006717233.1| S6 family IgA-specific metalloendopeptidase/...	61	3.00E-07	2	PF03224	V-ATPaseHN	56.4	1.6	0.00011	26.4	34	52
27	ref|WP_006717233.1| S6 family IgA-specific metalloendopeptidase/...	57	8.00E-06	n.d.								54
28	ref|WP_006717233.1| S6 family IgA-specific metalloendopeptidase/...	55	3.00E-05	5	PF13371	TPR_9	27.8	9.5	0.00064	17.1	20	55
29	ref|XP_005533922.1| PREDICTED: LOW QUALITY PROTEIN: staphylococc...	52	1.00E-04	20	PF12348	CLASP_N	22.5	14	0.00093	20.2	40	53
30	ref|XP_004688608.1| PREDICTED: laminin subunit gamma-1 [Condylur...	50	6.00E-04	21	PF13174	TPR_6	24.4	12	0.00081	14.4	16	55
31	n.d.			5	PF13174	TPR_6	70.8	0.56	3.80E-05	18	23	41
32	ref|WP_006717233.1| S6 family IgA-specific metalloendopeptidase/...	50	4.00E-04	8	PF13174	TPR_6	59.6	1.3	8.90E-05	16.8	20	41
33	n.d.			9	PF13174	TPR_6	37.6	5.1	0.00034	14.9	19	41
34	n.d.			n.d.								41
35	n.d.			n.d.								41
36	n.d.			27	PF12862	Apc5	37.1	5.3	0.00035	19.4	15	41
37	n.d.			4	PF13606	Ank_3	70.7	0.57	3.80E-05	19.8	14	41
38	n.d.			14	PF00023	Ank	30.7	7.8	0.00053	15.3	16	41
39	ref|NP_908998.1| retrotransposon-like protein 1 [Mus musculus] >...	51	2.00E-04	3	PF13812	PPR_3	73.8	0.43	2.90E-05	18.7	17	42
40	n.d.			76	PF10516	SHNi-TPR	35.7	5.7	0.00039	18.1	14	43
41	n.d.			n.d.								57
42	n.d.			5	PF13374	TPR_10	50.1	2.4	0.00016	17.4	17	57
43	ref|XP_965543.1| hypothetical protein NCU01912 [Neurospora crass...	57	1.00E-05	n.d.								57
44	ref|YP_001389869.1| cell wall-associated hydrolase [Clostridium ...	50	4.00E-04	16	PF13646	HEAT_2	24.3	12	0.00081	15.9	18	35
45	n.d.			n.d.								35
46	ref|YP_005677090.1| N-acetylmuramoyl-L-alanine amidase [Clostrid...	52	1.00E-04	1	PF14852	Fis1_TPR_N	90.6	0.019	1.30E-06	26.7	26	39
47	n.d.			11	PF14852	Fis1_TPR_N	57.8	1.5	0.0001	19.8	21	39
48	n.d.			7	PF13041	PPR_2	71.5	0.53	3.60E-05	20.4	22	40
49	ref|XP_003872552.1| conserved hypothetical protein [Leishmania m...	56	8.00E-06	1	PF12862	Apc5	80.5	0.2	1.30E-05	25	30	40
50	n.d.			4	PF13525	YfiO	36.5	5.4	0.00037	21.7	29	43
51	n.d.			1	PF13174	TPR_6	67.4	0.75	5.10E-05	17.7	17	43
52	n.d.			2	PF13174	TPR_6	69.5	0.63	4.20E-05	18	17	43
53	n.d.			1	PF13174	TPR_6	69.1	0.65	4.40E-05	17.9	18	43
54	ref|XP_002544663.1| predicted protein [Uncinocarpus reesii 1704]...	51	2.00E-04	1	PF12862	Apc5	81.1	0.18	1.20E-05	25.4	27	43
55	gb|ETW54908.1| hypothetical protein PFUGPA_02731 [Plasmodium fal...	52	1.00E-04	7	PF13374	TPR_10	39.1	4.7	0.00031	15.8	16	46
56	n.d.			12	PF13374	TPR_10	35.4	5.8	0.00039	15.5	16	46
57	n.d.			5	PF12862	Apc5	32.2	7.1	0.00048	19.3	18	47
58	n.d.			4	PF03704	BTAD	31.6	7.4	0.0005	19.5	31	47
59	n.d.			1	PF12862	Apc5	81.7	0.17	1.10E-05	26.1	26	48
60	ref|XP_001350389.1| conserved Plasmodium protein, unknown functi...	49	8.00E-04	n.d.								33
61	n.d.			n.d.								33
62	ref|XP_003872552.1| conserved hypothetical protein [Leishmania m...	52	1.00E-04	1	PF12862	Apc5	79.1	0.23	1.60E-05	25	33	43
63	ref|XP_002594366.1| hypothetical protein BRAFLDRAFT_277633 [Bran...	52	2.00E-04	8	PF00637	Clathrin	59.6	1.3	8.90E-05	23.1	28	54
64	n.d.			4	PF13374	TPR_10	27.1	9.9	0.00067	15.3	18	57
65	n.d.			7	PF12862	Apc5	29.1	8.7	0.00058	19.7	15	57
66	ref|XP_005822570.1| hypothetical protein GUITHDRAFT_118293 [Guil...	55	2.00E-05	8	PF13374	TPR_10	27.6	9.6	0.00065	15.2	20	54
67	n.d.			2	PF12862	Apc5	52.6	2.1	0.00014	22.4	28	60
68	ref|XP_003769481.1| PREDICTED: laminin subunit alpha-2 [Sarcophi...	62	2.00E-07	1	PF13374	TPR_10	93.6	0.0032	2.10E-07	27.8	32	61
69	n.d.			4	PF13174	TPR_6	53.8	1.9	0.00013	17.5	17	62
70	n.d.			1	PF12862	Apc5	81.8	0.16	1.10E-05	26.3	33	50
71	n.d.			4	PF13606	Ank_3	26	11	0.00072	15.9	17	50
72	n.d.			3	PF12862	Apc5	67.8	0.72	4.90E-05	22.4	27	37
73	ref|YP_001642797.1| phage minor structural protein [Bacillus wei...	51	3.00E-04	1	PF13374	TPR_10	95.1	0.00045	3.00E-08	29.8	34	50
74	gb|EJW88262.1| laminin subunit gamma-1, partial [Wuchereria banc...	51	4.00E-04	4	PF07980	SusD	47.5	2.8	0.00019	23	18	55
75	n.d.			8	PF08238	Sel1	27.1	10	0.00067	15.3	13	51
76	gb|EXX64005.1| Sps1p [Rhizophagus irregularis DAOM 197198w] 50 8e-0	4		n.d.								55
77	ref|XP_003872552.1| conserved hypothetical protein [Leishmania m...	67	8.00E-09	n.d.								51
78	ref|XP_003872552.1| conserved hypothetical protein [Leishmania m...	67	8.00E-09	n.d.								51
79	n.d.			8	PF12348	CLASP_N	54.7	1.8	0.00012	24.3	26	56
80	ref|WP_006717233.1| S6 family IgA-specific metalloendopeptidase/...	55	3.00E-05	1	PF12862	Apc5	73.7	0.43	2.90E-05	25.1	31	57
81	ref|WP_006717233.1| S6 family IgA-specific metalloendopeptidase/...	56	2.00E-05	1	PF12862	Apc5	73.9	0.42	2.80E-05	25.2	31	57
82	ref|XP_003767551.1| PREDICTED: laminin subunit gamma-1, partial ...	54	8.00E-05	2	PF00515	TPR_1	72.8	0.47	3.20E-05	20.1	21	58
83	ref|YP_005168772.1| hypothetical protein DND132_2766 [Desulfovib...	50	8.00E-04	1	PF00515	TPR_1	73.9	0.42	2.80E-05	20.2	21	58

Best hits from Blast are shown in the left part of the table: n.d. indicates that no hit was found below the E value cutoff of 0.0001. The right side shows the best ranking repeat protein families in Pfam resulting from HHsearch. The final set of proteins validated by crystallography and/or small angle X-ray scattering is highlighted in grey.

Supplementary Discussion 4 | Structure determination remarks

Due to the presence of 6 cysteine residues in the native protein, the DHR5 structure was solved by sulfur single wavelength anomalous dispersion (S-SAD) using a dataset collected at 7235 eV. A search for 6 individual sulfur atoms in SHELXD gave many clear solutions that led to near complete autobuilding of a poly-alanine backbone in SHELXE, which was further elaborated using the Autobuild module of Phenix. Ultimately, the final model for DHR5 was in good agreement with the design target structure, despite our initial difficulties in phasing by molecular replacement. While the SAD data set was limited to 1.85 Å, the final model was refined against the original data set (1.25Å). Both data sets were deposited in the Protein Data Bank.

The asymmetric unit for DHR8 was found to contain 4 copies of DHR8. Although the overall structure of the 4 copies is similar, the electron density for the N-terminal helix from two of these monomers is weak, suggesting that these helices are partially disordered in the crystal. Indeed, crystal packing of these helices in the designed conformation would have led to significant steric overlap with one another. As the corresponding helices in the remaining two DHR8 monomers were well-ordered and essentially as designed, these fully ordered models were used for further analysis.

The dataset collected for DHR14 had a large non-origin Patterson peak at fractional coordinates (0.000, 0.217, 0.000), suggesting the presence of translational NCS. However, consideration of the apparent space group, unit cell parameters, and plausible solvent content strongly indicated the presence of a single copy of DHR14 in the asymmetric unit. Given the relatively low pitch of this helical design and the translational pseudosymmetry between the N- and C-terminal halves of the protein, we suspected that intramolecular pseudotranslational NCS might account for the observed Patterson peak. Ultimately, a molecular replacement solution was obtained using 4 of the 8 designed helices of DHR14, and this was sufficient to bootstrap autobuilding of the remaining backbone using SHELXE. In the final model, the helical axis of DHR14 is closely aligned with the crystallographic b axis, and pseudotranslational NCS between the N- and C-terminal repeats with a translation of ~21 Å is in good agreement with the observed fractional Patterson peak at ~0.22 along b.

Supplementary Table 5 | Summary crystallization

Protein	Concentration (mg/ml)	Cryoprotectant added	Crystallization condition	Beamline	Phasing method	Space group	resolution
DHR4	23	7.5% ethylene glycol	0.2 M Ammonium acetate, 0.1M acetate pH 4.6, 30% PEG 4000	ALS 8.3.1	Molecular replacement	$P2_12_12_1$	1.55
DHR5	23	17.5% ethylene glycol	1.0 M Lithium chloride, 0.1 M HEPES pH 7.0, 20% (w/v) PEG 6000	ALS 8.3.1	Molecular replacement	$P2_122_1$	1.25
DHR5	23	17.5% ethylene glycol	20% (w/v) PEG 6000, 1 M Lithium chloride, 0.1 M BisOTris pH 7.0	ALS 8.3.1	Sulfur SAD	$P2_122_1$	1.85
DHR7	19	none	0.2 M Sodium chloride, 0.1 M CHES pH 9.5, 50% (v/v) PEG 400	ALS 8.3.1	Molecular replacement	C121	2.60
DHR8	7	7.5% ethylene glycol	0.19 M Calcium chloride, 0.095 M HEPES pH 7.5, 26.6% (v/v) PEG 400, 5% (v/v) Glycerol	ALS 8.3.1	Molecular replacement	$P2_12_12$	1.80
DHR10	23	17.5% ethylene glycol	0.2 M Magnesium formate, 20% (w/v) PEG 3350	ALS 8.3.1	Molecular replacement	$P2_122_1$	1.20
DHR14	70	17.5% ethylene glycol	0.2 M Lithium nitrate, 20% (w/v) PEG 3350	ALS 8.3.1	Molecular replacement	$C222_1$	1.30
DHR18	88	none	0.1 M PhosphateOcitrate pH 4.2, 40% (v/v) PEG 300	ALS 8.3.1	Molecular replacement	P1	1.75
DHR49	31	15% ethylene glycol	0.2 M diOPotassium hydrogen phosphate, 20% (w/v) PEG 3350	ALS 8.3.1	Molecular replacement	$I4_1$	1.70
DHR53	63	12.5% ethylene glycol	26% (w/v) PEG 6000, 0.1 M Sodium citrate, pH 2.2	ALS 8.3.1	Molecular replacement	$C12_1$	1.90
DHR54	92	none	0.2 M Sodium chloride, 0.1 M Na/K phosphate pH 6.2, 50% (v/v) PEG 200	ALS 8.3.1	Molecular replacement	$P2_12_12_1$	1.50
DHR64	43	none	0.1 M Sodium acetate pH 4.5, 35% (v/v) MPD	ALS 8.3.1	Molecular replacement	P2	2.50
DHR64	43	25% ethylene glycol	0.1 M Sodium acetate pH 5.0, 10% (v/v) MPD	ALS 8.3.1	Molecular replacement	P6	2.90
DHR71	147	none	0.1 M CHES pH 9.5, 50% (v/v) PEG 200	ALS 8.3.1	Molecular replacement	$P2_12_12_1$	1.70
DHR76	66	30% ethylene glycol	0.1 M MES pH 5.0, 5% (w/v) PEG 6000	ALS 8.3.1	Molecular replacement	$P2_12_12_1$	3.35
DHR79	22	35% ethylene glycol	1.6 M Sodium citrate pH 6.5	ALS 8.3.1	Molecular replacement	$P4_32_12$	1.90
DHR81	39	15% ethylene glycol	0.08 M Sodium acetate pH 4.6, 20% (v/v) Glycerol	ALS 8.2.1	Molecular replacement	I23	2.05

Supplementary Table 6. Data collection and refinement statistics for DHR4 and DHR5

Data collection	DHR4	DHR5
Beamline	ALS 8.3.1	ALS 8.3.1
Wavelength (Å)	1.12	1.12
Space group	$P2_12_12_1$	$P2_122_1$
Unit cell parameters (Å, °)	a = 43.8, b = 56.2, c =66.4 Å $\alpha = \beta = \gamma = 90.0°$	a = 47.3, b = 51.1, c =72.9 Å $\alpha = \beta = \gamma = 90.0°$
Resolution (Å)	50 - 1.55 (1.59-1.55) [a]	50 - 1.25 (1.28-1.25) [a]
Observations	172,015	334,858
Unique Reflections	24,187 (1,729) [a]	47,733 (2,877) [a]
Redundancy	7.1 (7.2) [a]	7.0 (5.7) [a]
Completeness (%)	98.7 (97.6) [a]	96.0 (79.6) [a]
$<I/\sigma_I>$	16.6 (1.0) [a]	18.2 (1.2) [a]
CC1/2	0.99 (0.42)	1.00 (0.65)
R_{sym} [b]	0.07 (1.98) [a]	0.05 (1.53) [a]
Refinement statistics		
Resolution (Å)	50 – 1.55	50 – 1.25
Reflections (work)	22,937	44,703
Refections (test)	1,207	1,758
$R_{cryst}(\%)$[c] / $R_{free}(\%)$[d]	19.1 / 22.0	17.5 / 19.4
Average B ($Å^2$)	35.1	26.6
Wilson B ($Å^2$)	22.5	13.8
Protein atoms	3,667	3725
Waters	91	191
Other	0	0
RMSD from ideal geometry		
Bond length (Å)	0.005	0.005
Bond angles (°)	0.77	0.84
Ramachandran statistics (%)[e]		
Favored	99.1	99.5
Outliers	0.0	0.0
PDB Code	5CWB	5CWC

Supplementary Table 7. Data collection and refinement statistics for DHR7 and DHR8

Data collection	DHR7	DHR8
Beamline	ALS 8.3.1	ALS 8.3.1
Wavelength (Å)	1.12	1.12
Space group	C2	$P2_12_12$
Unit cell parameters (Å, °)	a = 84.7, b = 29.3, c = 63.2 Å $\alpha = \gamma = 90.0°$, $\beta = 123.5$	a = 75.7, b = 76.4, c =33.1 Å $\alpha = \beta = \gamma = 90.0°$
Resolution (Å)	50 - 2.60 (2.67-2.60) [a]	50 - 1.80 (1.85-1.80) [a]
Observations	13,960	512,567
Unique Reflections	4,124 (322) [a]	72,203 (5,260) [a]
Redundancy	3.4 (3.4) [a]	7.1 (7.2) [a]
Completeness (%)	98.3 (97.9) [a]	99.9 (100.0) [a]
$<I/\sigma_I>$	4.9 (1.2) [a]	12.1 (1.1) [a]
CC1/2	0.99 (0.63)	1.00 (0.48)
R_{sym} [b]	0.15 (0.81) [a]	0.09 (1.85) [a]
Refinement statistics		
Resolution (Å)	50 – 2.60	50 – 1.80
Reflections (work)	3,916	68,562
Refections (test)	206	3,591
R_{cryst}(%)[c] / R_{free}(%)[d]	25.3 / 30.9	20.5 / 22.8
Average B ($Å^2$)	66.5	44.8
Wilson B ($Å^2$)	55.6	27.7
Protein atoms	2,521	11,104
Waters	0	466
Other	0	31
1GRMSD from ideal geometry		
Bond length (Å)	0.004	0.005
Bond angles (°)	0.58	0.73
Ramachandran statistics (%)[e]		
Favored	99.4	99.3
Outliers	0.0	0.0
PDB Code	5CWD	5CWF

Supplementary Table 8. Data collection and refinement statistics for DHR10 and DHR14

Data collection	DHR10	DHR14
Beamline	ALS 8.3.1	ALS 8.3.1
Wavelength (Å)	1.12	1.12
Space group	$P2_122_1$	$C222_1$
Unit cell parameters (Å, °)	a = 47.6, b = 50.6, c = 72.4 Å α = β = γ = 90.0°	a = 41.0, b = 104.4, c = 90.2 Å α = β = γ = 90.0°
Resolution (Å)	50 - 1.20 (1.23-1.20) [a]	50 - 1.30 (1.33-1.30) [a]
Observations	313,175	964,864
Unique Reflections	54,627 (3,516) [a]	47,853 (3,495) [a]
Redundancy	5.6 (2.8) [a]	20.2 (17.0) [a]
Completeness (%)	98.5 (87.4) [a]	99.7 (99.8) [a]
$\langle I/\sigma_I \rangle$	22.7 (1.9) [a]	19.2 (1.5) [a]
CC1/2	1.00 (0.81)	1.00 (0.85)
R_{sym} [b]	0.04 (0.53) [a]	0.08 (2.51) [a]
Refinement statistics		
Resolution (Å)	50 – 1.20	50 – 1.30
Reflections (work)	51,813	45,170
Refections (test)	2,730	2,377
R_{cryst}(%)[c] / R_{free}(%)[d]	14.1 / 16.9	16.2 / 18.6
Average B (Å^2)	21.3	30.1
Wilson B (Å^2)	12.0	15.2
Protein atoms	3,839	2,860
Waters	238	240
Other	10	20
1GRMSD from ideal geometry		
Bond length (Å)	0.007	0.018
Bond angles (°)	1.07	1.52
Ramachandran statistics (%)[e]		
Favored	99.6	100.0
Outliers	0.0	0.0
PDB Code	5CWG	5CWH

Supplementary Table 9. Data collection and refinement statistics for DHR18 and DHR49

Data collection	DHR18	DHR49
Beamline	ALS 8.3.1	ALS 8.3.1
Wavelength (Å)	1.12	1.12
Space group	P1	P1
Unit cell parameters (Å, °)	a = 24.9, b = 43.7, c = 48.8 α = 116.6, β = 90.7, γ = 102.2	a = 46.7, b = 61.7, c = 61.7 α = 81.9, β = 67.8, γ = 67.9
Resolution (Å)	50 – 1.75 (1.80-1.75) [a]	50 - 1.80 (1.85-1.80) [a]
Observations	36,231	103,465
Unique Reflections	17,261 (1,247) [a]	52,418 (3,824) [a]
Redundancy	2.1 (2.1) [a]	2.0 (2.0) [a]
Completeness (%)	94.8 (94.3) [a]	95.6 (94.5) [a]
$\langle I/\sigma_I \rangle$	9.8 (1.1) [a]	9.9 (1.4) [a]
CC1/2	1.00 (0.54)	1.00 (0.57)
R_{sym} [b]	0.05 (0.85) [a]	0.04 (0.59) [a]
Refinement statistics		
Resolution (Å)	50 – 1.75	50 – 1.80
Reflections (work)	16,378	47,180
Refections (test)	866	1,864
R_{cryst}(%)[c] / R_{free}(%)[d]	18.5 / 22.2	24.0 / 28.8
Average B (Å^2)	41.2	49.4
Wilson B (Å^2)	23.9	28.9
Protein atoms	3,757	10,044
Waters	69	370
Other	5	0
1GRMSD from ideal geometry		
Bond length (Å)	0.010	0.005
Bond angles (°)	1.01	0.72
Ramachandran statistics (%)[e]		
Favored	99.2	98.7
Outliers	0.4	0.0
PDB Code	5CWI	5CWJ

Supplementary Table 10. Data collection and refinement statistics for DHR53 and DHR54

Data collection	DHR53	DHR54
Beamline	ALS 8.3.1	ALS 8.3.1
Wavelength (Å)	1.12	1.12
Space group	C2	$P2_12_12_1$
Unit cell parameters (Å, °)	a = 127.4, b = 24.4, c = 60.5 $\alpha = \gamma = 90.0, \beta = 116.9$	a = 55.3, b = 69.1, c = 82.6 $\alpha = \beta = \gamma = 90.0$
Resolution (Å)	50 – 1.90 (1.95-1.90) [a]	50 - 1.50 (1.54-1.50) [a]
Observations	46,741	360,741
Unique Reflections	13,327 (949) [a]	51,178 (3,720) [a]
Redundancy	3.5 (3.6) [a]	7.0 (7.1) [a]
Completeness (%)	98.3 (99.2) [a]	99.6 (99.9) [a]
$<I/\sigma_I>$	13.2 (1.0) [a]	15.1 (1.0) [a]
CC1/2	1.00 (0.74)	1.00 (0.49)
R_{sym} [b]	0.05 (1.67) [a]	0.06 (1.80) [a]
Refinement statistics		
Resolution (Å)	50 – 1.90	50 – 1.50
Reflections (work)	12,581	45,069
Refections (test)	663	1,817
R_{cryst}(%)[c] / R_{free}(%)[d]	24.5 / 28.5	18.3 / 22.2
Average B ($Å^2$)	83.9	43.1
Wilson B ($Å^2$)	38.0	23.2
Protein atoms	2,755	5,515
Waters	46	168
Other	0	0
1GRMSD from ideal geometry		
Bond length (Å)	0.006	0.011
Bond angles (°)	0.81	1.25
Ramachandran statistics (%)[e]		
Favored	97.7	98.9
Outliers	0.6	0.3
PDB Code	5CWK	5CWL

Supplementary Table 11. Data collection and refinement statistics for DHR64 and DHR71

Data collection	DHR64	DHR71
Beamline	ALS 8.3.1	ALS 8.3.1
Wavelength (Å)	1.12	1.12
Space group	$P6_5$	$P2_12_12_1$
Unit cell parameters (Å, °)	a = b = 98.3, c = 48.6 α = β = 90.0, γ = 120.0	a = 23.3, b = 82.9, c = 103.3 α = β = γ = 90.0
Resolution (Å)	50 - 2.90 (2.98-2.90) [a]	50 - 1.70 (1.74-1.70) [a]
Observations	67,571	158,634
Unique Reflections	6,086 (448) [a]	22,958 (1,666) [a]
Redundancy	11.1 (11.3) [a]	6.9 (7.1) [a]
Completeness (%)	100.0 (100.0) [a]	99.6 (100.0) [a]
$\langle I/\sigma_I \rangle$	14.4 (1.3) [a]	15.0 (1.3) [a]
CC1/2	1.00 (0.44)	1.00 (0.64)
R_{sym} [b]	0.15 (2.14) [a]	0.08 (1.64) [a]
Refinement statistics		
Resolution (Å)	50 – 2.90	50 – 1.70
Reflections (work)	5,268	21,747
Refections (test)	277	1,144
R_{cryst}(%)[c] / R_{free}(%)[d]	21.4 / 26.5	20.6 / 25.0
Average B (Å^2)	104.9	36.1
Wilson B (Å^2)	72.9	24.4
Protein atoms	3,656	3,257
Waters	29	95
Other	0	0
1GRMSD from ideal geometry		
Bond length (Å)	0.019	0.010
Bond angles (°)	1.70	1.14
Ramachandran statistics (%)[e]		
Favored	93.3	99.0
Outliers	0.9	0.0
PDB Code	5CWM	5CWN

Supplementary Table 12. Data collection and refinement statistics for DHR76 and DHR79

Data collection	DHR76	DHR79
Beamline	ALS 8.3.1	ALS 8.3.1
Wavelength (Å)	1.12	1.12
Space group	$P2_12_12_1$	$P4_32_12$
Unit cell parameters (Å, °)	a = 62.5, b = 66.3, c = 102.2 $\alpha = \beta = \gamma = 90.0°$	a = b = 94.3, c = 77.6 Å $\alpha = \beta = \gamma = 90.0°$
Resolution (Å)	50 – 3.55 (3.44-3.55) [a]	50 - 1.90 (1.95-1.90) [a]
Observations	45,810	398,338
Unique Reflections	6,514 (470) [a]	28,192 (2,042) [a]
Redundancy	7.0 (7.2) [a]	14.1 (14.4) [a]
Completeness (%)	99.8 (100.0) [a]	100.0 (100.0) [a]
$<I/\sigma_I>$	10.9 (1.2) [a]	17.5 (1.1) [a]
CC1/2	1.00 (0.33)	1.00 (0.48)
R_{sym} [b]	0.18 (1.91) [a]	0.11 (3.61) [a]
Refinement statistics		
Resolution (Å)	50 – 3.55	50 – 1.90
Reflections (work)	6,162	26,730
Refections (test)	322	1,406
R_{cryst}(%)[c] / R_{free}(%)[d]	37.2 / 40.8	19.3 / 23.8
Average B (Å^2)	242.9	58.6
Wilson B (Å^2)	111.5	34.2
Protein atoms	6,992	3,591
Waters	0	63
Other	0	0
1GRMSD from ideal geometry		
Bond length (Å)	0.004	0.013
Bond angles (°)	0.83	1.29
Ramachandran statistics (%)[e]		
Favored	98.6	98.3
Outliers	0.0	0.4
PDB Code	5CWO	5CWP

Supplementary Table 13. Data collection and refinement statistics for DHR81

Data collection	DHR81
Beamline	ALS 8.2.1
Wavelength (Å)	1.00
Space group	I23
Unit cell parameters (Å, °)	a = b = c = 120.0 $\alpha = \beta = \gamma = 90.0$
Resolution (Å)	50 - 2.05 (2.10-2.05) [a]
Observations	397,887
Unique Reflections	18,193 (1,317) [a]
Redundancy	21.9 (22.4) [a]
Completeness (%)	100.0 (100.0) [a]
$<I/\sigma_I>$	16.8 (1.1) [a]
CC1/2	1.00 (0.32)
R_{sym} [b]	0.31 (3.53) [a]
Refinement statistics	
Resolution (Å)	50 – 2.05
Reflections (work)	16,326
Refections (test)	860
R_{cryst}(%)[c] / R_{free}(%)[d]	21.7 / 26.6
Average B (Å^2)	49.3
Wilson B (Å^2)	31.4
Protein atoms	3,654
Waters	51
Other	14
1GRMSD from ideal geometry	
Bond length (Å)	0.007
Bond angles (°)	0.85
Ramachandran statistics (%)[e]	
Favored	98.3
Outliers	0.4
PDB Code	5CWQ

Supplementary Table 14. Data collection and phasing statistics for DHR5 (S-SAD)

Data collection	
Beamline	ALS 8.3.1
Space group	$P2_122_1$
Unit cell parameters (Å, °)	a = 47.3, b = 50.7, c = 73.0 $\alpha = \beta = \gamma = 90.0$
Wavelength (Å)	1.714
Resolution (Å)	50 – 1.85 (1.90-1.85) [a]
Observations	556,066
Unique Reflections	28,741 (1,966) [a]
Redundancy	19.3 (7.2) [a]
Completeness (%)	99.2 (91.0) [a]
$\langle I/\sigma_I \rangle$	31.9 (3.2) [a]
CC1/2	1.00 (0.92)
R_{sym} [b]	0.08 (0.50) [a]
FOM	0.58
PDB Code	5CWC

Supplementary Tables 6-14 | Crystallography data collection

[a] Numbers in parentheses refer to the highest resolution shell.

[b] $R_{sym} = \Sigma_{hkl} \Sigma_i | I_{hkl,i} - \langle I_{hkl} \rangle | / \Sigma_{hkl} \Sigma_i I_{hkl,I}$, where $I_{hkl,i}$ is the scaled intensity of the i^{th} measurement of relection h, k, l, $\langle I_{hkl} \rangle$ is the average intensity for that reflection, and n is the redundancy [24].

[c] $R_{cryst} = \Sigma_{hkl} | F_o - F_c | / \Sigma_{hkl} | F_o |$ x 100

[d] R_{free} was calculated as for R_{cryst}, but on a test set comprising 5% of the data excluded from refinement.

[e] Calculated using Molprobity[25].

Supplementary Discussion 5 | Small Angle X-ray Scattering (SAXS) analysis

Guinier and P(r) analysis were done using using ATSAS [26]. The Porod exponent was determined from a linear regression analysis (I vs q) of the top of the first peak in the Porod-Debye plot (q^4*I(q) vs q^4) of the scattering data, implemented in SCÅTTER, available at beamline 12.3.1 [27,28]. The molecular mass in solution was calculated using SCÅTTER [29]. The results are summarized in Supplementary Table 15.

25% of the designs had molecular weights in solution that were significantly greater than the predicted molecular weight (1.2-4 fold), suggesting that these designs formed multimeric assemblies or a small portion of aggregates[29]. All 55 designs had Porod exponents (P_E) greater than 2.9, indicating significant levels of folded protein; 67% of the designs had a P_E of 3.4-4, indicating a well-folded core[28]. Of the 15 proteins that crystallized, the majority (66%) had P_E of 3.9-4, consistent with more well-packed proteins being easier to crystallize.

Radius of gyration (Rg) and maximum of distance distribution (dmax) were calculated from real space distance distribution P(r). Distance distribution comparisons between experimental data and models are included in supporting_experimental_data.pdf. Among the models confirmed by crystallography, DHR 49 and 76 formed dimers in solution. The experimental data were fit using models based on the dimer configuration observed in the crystal structure. DHR 5 tendency to aggregation (see SEC in supporting_experimental_data.pdf) affected the SAXS profile resulting in a high Molecular weight and Vr above our acceptance threshold.

If molecular mass and Rg of models were within a 25% error from experimental data and Vr was below 2.5, the models were considered able to recapture the SAXS data. Dmax errors are generally within 25%; higher values are highlighted in Supplementary Table 15.

43 designs satisfied our requirements: DHR 1 2 3 4 7 8 9 10 14 15 18 20 21 23 24 26 27 31 32 36 39 46 47 49 52 53 54 55 57 58 59 62 64 68 70 71 72 76 77 78 79 80 81 82. The corresponding models, including also DHR5, are shown in Extended Data Fig. 9.

Supplementary Table 15 | Summary of SAXS analysis

		MW				q range		Guinier				gnom p(r)			Model p(r)		Model Fit		Crystal p(r)		Crystal Fit		
DHR	protein length	expected	SAXS	diff w/in 25%	Porod	q min	q end	start range	end range	Rg	I0	dmax	Rg	I0	dmax	Rg	Vr 0.015<q q<0.25	chi	dmax	Rg	Vr 0.015<q q<0.25	chi	accepted
1	187	19.983	23.5	YES	4	0.02767	0.324398	0.02767	0.059292	23	1078	90	24.46	1098	65	19.45	2.02	6.62					yes*
3	179	19.949	21.8	YES	4	0.032535	0.325006	0.032535	0.05686	22.8	441	80	23.2	438	65	18.91	1.53	4.13					yes
4	211	23.8	21.8	YES	4	0.038616	0.322574	0.038616	0.074495	18.5	686.32	55	18.44	689	55	18.69	1.76	2.44	54	18.5	1.46	2.21	yes
5	210	23.714	73.5	NO	3.5	0.013075	0.328046										3.65	2.04					no
6	178	19.983	31.8	NO	3.7	0.013075	0.328019										2.87	27.01					no
7	179	19.066	16.6	YES	3.7	0.034359	0.325006	0.034359	0.070847	19.2	209.98	70	19.8	211	60	18.72	1.04	1.73	65	18.63	1.23	1.93	yes
8	191	21.339	21.3	YES	4	0.027059	0.326803	0.027059	0.063539	21.6	511.1	75	22.1	513	72	20.82	2.29	1.58	70	21.4	1.76	0.99	yes
9	211	23.734	22.8	YES	4	0.024021	0.322546	0.024021	0.05017	18.2	538.67	55	18.5	546	52	18.32	2.01	1.51					yes
10	210	23.605	26.5	YES	4	0.022805	0.322574	0.022805	0.05017	24.3	600.99	90	25.02	603	62	20.37	1.77	5.00	76	22.57	1.51	2.67	yes
11	187	20.63	78.7	NO	3.4	0.013075	0.328046										6.63	10.10					no
13	156	17.017	16.6	YES	2.8	0.028278	0.320143	0.028278	0.05686	24	81.84	120	29.28	89.3	52	17.83	5.03	4.69					no
14	172	19.552	19.3	YES	3.9	0.016724	0.321967	0.016724	0.071455	19	244.21	58	19.2	245	65	19.59	0.99	2.12	58	19.04	0.75	1.42	yes
15	196	22.58	22.9	YES	4	0.020372	0.319535	0.020372	0.070238	19.4	199.2	65	19.45	200	63	19.42	1.28	1.69					yes
16	188	21.495	39.6	NO	3.9	0.02767	0.323182										4.32	22.02					no
17	228	25.733	30	YES	3.1	0.024021	0.320751	0.024021	0.0374	35.2	78.164	150	41.52	83.1	69	21.09	4.11	4.93					no
18	250	26.993	22.3	YES	2.9	0.012467	0.321967	0.012467	0.039224	25	1236.1	80	23.1	1187	76	20.67	0.88	6.81	81	21.5	0.82	4.91	yes
19	163	18.64	25.8	NO	3.9	0.024629	0.323182										4.67	3.15					no
20	163	18.284	17.6	YES	4	0.020372	0.320751	0.020372	0.071455	17.7	287.32	50	17.73	289	51	17.75	1.18	3.06					yes
21	164	18.765	20.4	YES	3.9	0.02098	0.325614	0.02098	0.06659	20.6	2068.3	80	21.4	209	68	19.96	0.74	2.26					yes
23	171	19.974	21.4	YES	4	0.024627	0.326195	0.024627	0.061107	22	1625	92	23.77	1667	80	20.98	1.66	3.83					yes
24	195	22.344	24.5	YES	4	0.016115	0.321359	0.016115	0.056252	22	396.79	80	22.6	399	70	21.52	0.93	3.70					yes
25	191	21.781	38.8	NO	4	0.010034	0.322574										4.02	18.21					no
26	219	25.366	25.5	YES	4	0.022195	0.327411	0.022195	0.063539	21.3	737.2	90	22.2	740	60	19.93	1.21	2.29					yes
27	227	26.734	24.8	YES	4	0.016724	0.317711	0.016724	0.067806	20.3	215.01	60	21.8	215	58	19.74	1.27	1.67					yes
31	175	19.745	19	YES	3.6	0.024627	0.319507	0.024627	0.065971	20.8	1652	84	21.1	1688	70	21.73	0.68	1.29					yes
32	175	19.756	19	YES	3.4	0.028278	0.320143	0.028278	0.065374	20.7	633.1	72	21.6	639	70	21.97	1.80	0.64					yes
36	175	19.813	20.3	YES	3.1	0.03071	0.327438	0.03071	0.053211	24.9	504.19	100	27.5	524	80	24.56	1.47	0.86					yes
39	179	20.141	19	YES	3.2	0.022195	0.327411	0.022195	0.061107	22	980.98	92	23.58	1003	67	21.35	0.83	2.42					yes
46	167	19.28	16.2	YES	3.1	0.032531	0.327411	0.032531	0.068403	20.2	2067.3	85	21.7	2121	79	20.67	1.94	14.10					yes
47	167	19.395	16.6	YES	3.2	0.022805	0.32683	0.022805	0.0599	21.1	1063.8	85	22.25	1080	76	20.34	1.81	5.13					yes
49	171	39.122	35.3	YES	4	0.013075	0.328046	0.014291	0.061117	22.4	5917.1	94	22.34	5922	70	22.31	0.77	2.84	67	22.41	0.93	4.65	yes
51	183	20.432	31.5	NO	3.6	0.011859	0.323763										2.89	9.63					no
52	183	20.381	26.5	YES	3.9	0.013075	0.328046	0.013075	0.056252	23.7	1545.7	85	24.71	1568	66	20.48	2.33	6.40					yes*
53	183	20.467	16.6	YES	3.3	0.011859	0.323763	0.011859	0.035571	25	7352.8	85	24.07	7122	75	21.85	0.74	11.44	87	21.18	0.82	14.05	yes
54	183	20.266	22.1	YES	4	0.025845	0.322574	0.025845	0.06659	20.6	2428.3	75	20.99	2443	57	19.26	1.06	4.32	65	20.1	0.93	2.71	yes
55	195	21.56	28.1	YES	4	0.015507	0.323763	0.015507	0.39219	25.08	10530	100	25.9	10410	65	20.67	1.52	30.66					yes
56	195	21.731	28.1	YES	4	0.017939	0.328019	0.017939	0.045299	27.2	3507.5	100	27.73	3501	90	27.14	2.69	24.89					no
57	199	22.3	24.9	YES	3.6	0.033143	0.317103	0.033143	0.065374	21.1	5512	63	21.5	5596	67	21.03	0.71	3.89					yes
58	199	22.318	16.9	YES	3.4	0.019764	0.320143	0.019764	0.067198	20.4	610.87	70	21	617	72	22.04	1.36	2.70					yes
59	203	22.459	19.6	YES	3.9	0.033139	0.321331	0.033139	0.059891	22.3	2647	95	24.22	2723	70	20.41	1.96	11.67					yes
62	183	20.044	16.4	YES	3.3	0.32535	0.320143	0.32535	0.070238	19.5	582.67	70	20	589	65	19.42	1.01	1.12					yes
64	239	27.18	26.2	YES	3	0.025237	0.323182	0.025237	0.061117	22.5	847.94	76	23.1	858	67	22.12	1.81	1.35	66	21.59	1.71	1.39	yes
66	227	25.153	25.4	YES	4	0.02767	0.323182	0.02767	0.063549	21.5	568.3	92	22.85	582.8	64	20.39	2.51	1.15					no
67	251	28.663	29.4	YES	2.8	0.02949	0.32133	0.02949	0.050162	26.9	275.86	110	29.74	289.6	86	23.83	3.71	2.74					no
68	255	29.804	28.4	YES	3.3	0.022803	0.327411	0.022803	0.059891	22.9	2460	80	23.34	2483	75	23.15	0.71	2.48					yes
70	211	23.8	18.3	YES	3	0.028275	0.327411	0.028275	0.056243	22.9	1463.4	90	24.1	1486	76	21.86	1.36	8.49					yes
71	211	23.566	25.8	YES	3.2	0.022197	0.325006	0.022197	0.056252	24.2	3659.3	90	25.14	3701	78	23.26	1.00	3.31	71	23.13	0.95	3.11	yes
72	159	17.341	15.4	YES	3	0.033139	0.327411	0.033139	0.058675	24.6	685.86	90	23.45	707	67	20.46	1.59	1.83					yes
76	231	52.526	57.6	YES	4	0.01125	0.323762	0.020978	0.04773	28.8	7222.7	110	29.79	7298	79	26.38	1.44	5.21	79	26.31	1.65	5.44	yes
77	215	24.919	22	YES	3.3	0.025842	0.323154	0.025842	0.06597	20.4	2032.8	65	20.89	2043	62	20.54	0.70	1.95					yes
78	215	24.853	22.7	YES	3.2	0.025237	0.319535	0.025237	0.065982	20	1451.5	65	21.1	1458	62	20.75	1.24	1.04					yes
79	235	26.366	24.6	YES	4	0.022194	0.317074	0.022194	0.06293	20	2861.2	58	20.46	2918	63	20.83	0.88	3.77	62	20.72	0.63	3.05	yes
80	239	26.805	24.7	YES	4	0.031314	0.322546	0.031314	0.059282	19.9	2330.4	57	19.9	2384	64	20.04	0.90	2.72					yes
81	239	26.888	23.7	YES	4	0.022805	0.324398	0.022805	0.065374	19.4	2239.6	55	19.6	2267	60	19.61	0.81	1.65	64	20.11	0.89	3.17	yes
82	243	27.434	24.5	YES	3.3	0.024018	0.322546	0.024018	0.06293	21.7	2106.3	70	21.9	2111	70	21.24	0.64	3.47					yes

Vr > 2.5

difference in Rg or Dmax between model and SAXS data > 25%

* designs with dmax discrepancy > 25% and 2 < Vr < 2.5

When experimental mass deviated from expected mass by more than 25% (diff w/in 25% = NO), the design was not further analyzed by Guinier or gnom. Vr is always above 2.5 in these cases.

Supplementary Table 16 | Protein Sequences

name	sequence
DHR1	MGCDQVAKDASSTIREVIEKNPNYSEKVADVAAKIVKKIIEGNPNGCDCVAKAASSIIRAVIEKNPNYSEVVADVAAAIVKAIIEGNPN GCDCVAKAASSIIRAVIEKNPNYSEVVADVAAAIVKAIIEGNPNGRDCVRKAASSIIRAVQEKNPNYSEVVEDVKRAIEKAIKEGNPNG WLEHHHHHH
DHR2	MSDADEAAKEANKAENKARNRNDDEAAKAVKLIKEAIERAKKRNESDAVEAAKEAAKALNKALNRNDDEAAKAVALIAEAIIRALKRNE SDAVEAAKEAAKALNKALNRNDDEAAKAVALIAEAIIRALKRNESDAVEKAKEAAKNLNKALNRNDDEQAKHVAKQAENIIRALKRNES WLEHHHHH
DHR3	MSSEDTVRKIAQKCSEAIRESNDCEEAARKCAKTISEAIRESNSSELAVRIIAQVCSEAIRESNDCECAARICAKIISEAIRESNSSEL AVRIIAQVCSEAIRESNDCECAARICAKIISEAIRESNSSELAKRIIKQVCSEAKRESNDTECAKRICTKIKSEAKRESNSWLEHHHHH H
DHR4	MSYEDECEEKARRVAEKVERLKRSGTSEDEIAEEVAREISEVIRTLKESGSSYEVICECVARIVAEIVEALKRSGTSEDEIAEIVARVI SEVIRTLKESGSSYEVICECVARIVAEIVEALKRSGTSEDEIAEIVARVISEVIRTLKESGSSYEVIKECVQRIVEEIVEALKRSGTSE DEINEIVRRVKSEVERTLKESGSSWLEHHHHHH
DHR5	MSSEKEELRERLVKICVENAKRKGDDTEEAREAAREAFELVREAAERAGIDSSEVLELAIRLIKECVENAQREGYDISEACRAAAEAFK RVAEAAKRAGITSSEVLELAIRLIKECVENAQREGYDISEACRAAAEAFKRVAEAAKRAGITSSETLKRAIEEIRKRVEEAQREGNDIS EACRQAAEEFRKKAEELKRRGDGWLEHHHHHH
DHR6	MSEEKEEALKKVREAAKKLGSSDEEARKCFEEAREWAERTGSSAYEAAEALFKVLEAAYKLGSSAEEACECFNQAAEWAERTGSGAYEA AEALFKVLEAAYKLGSSAEEACECFNQAAEWAERTGSGAYEAAERLFEELERAYEEGSSAEEACEEFNKKEEEAHRKGKKWLEHHHHHH
DHR7	MSTKEDARSTCEKAARKAAESNDEEVAKQAAKDCLEVAKQAGMPTKEAARSFCEAAARAAAESNDEEVAKIAAKACLEVAKQAGMPTKE AARSFCEAAARAAAESNDEEVAKIAAKACLEVAKQAGMPTKEAARSFCEAAKRAAKESNDEEVEKIAKKACKEVAKQAGMPWLEHHHHH H
DHR8	MSDEMKKVMEALKKAVELAKKNNDDEVAREIERAAKEIVEALRENNSDEMAKVMLALAKAVLLAAKNNDDEVAREIARAAAEIVEALRE NNSDEMAKVMLALAKAVLLAAKNNDDEVAREIARAAAEIVEALRENNSDEMAKKMLELAKRVLDAAKNNDDETAREIARQAAEEVEADR ENNSWLEHHHHHH
DHR9	MSYEDEAEEKARRVAEKVERLKRSGTSEDEIAEEVAREISEVIRTLKESGSSYEVIAEIVARIVAEIVEALKRSGTSEDEIAEIVARVI SEVIRTLKESGSSYEVIAEIVARIVAEIVEALKRSGTSEDEIAEIVARVISEVIRTLKESGSSYEVIKEIVQRIVEEIVEALKRSGTSE DEINEIVRRVKSEVERTLKESGSSWLEHHHHHH
DHR10	MSSEKEELRERLVKIVVENAKRKGDDTEEAREAAREAFELVREAAERAGIDSSEVLELAIRLIKEVVENAQREGYDISEAARAAAEAFK RVAEAAKRAGITSSEVLELAIRLIKEVVENAQREGYDISEAARAAAEAFKRVAEAAKRAGITSSETLKRAIEEIRKRVEEAQREGNDIS EAARQAAEEFRKKAEELKRRGDGWLEHHHHHH
DHR11	MSDADEAAKEANKAENKARNRNDDEAAKAVKLCKEAIERAKKRNESDAVEAAKEAAKALNKALNRNDDEAAKAVALCCEAIIRALKRNE SDAVEAAKEAAKALNKALNRNDDEAAKAVALCCEAIIRALKRNESDAVEKAKEAAKNLNKALNRNDDEQAKHVAKQCENIIRALKRNES WLEHHHHHH
DHR12	MDDEEQCREIAEKAKQTYTDDEEIARIIAEAARQTTTDDEEICRCIAEAAKQTYTDDEEIARIIAYAARQTTTDDEEICRCIAEAAKQT YTDDEEIARIIAYAARQTTTDDEEIERCIEEAAKQTYTDDEEIERIKEYARRQTTTDGWLEHHHHHH
DHR13	MNAEDKAREVLKELKDEGSPEEEAARQVLKDLNREGSNAEDAARAVLKALKDEGSPEEEAARAVLKALNREGSNAEDAARAVLKALKDE GSPEEEAARAVLKALNREGSNEEDASRAVLKALKDEGSPEEEARRAVEKALNREGSNGWLEHHHHHH
DHR14	MDSEEVNERVKQLAEKAKEATDKEEVIEIVKELAELAKQSTDSELVNEIVKQLAEVAKEATDKELVIYIVKILAELAKQSTDSELVNEI VKQLAEVAKEATDKELVIYIVKILAELAKQSTDSELVNEIVKQLEEVAKEATDKELVEHIEKILEELKKQSTDGWLEHHHHHH
DHR15	MNDERQKQREEVRKLAEELASKATDEELIKEIKKCAQLAEELASRSTNDELIKQILEVAKLAFELASKATDEELIKEILKCCQLAFELA SRSTNDELIKQILEVAKLAFELASKATDEELIKEILKCCQLAFELASRSTNDEEIKQILETAKEAFERASKATDEEEIKEILKKCQEKF EKKSRSTNGWLEHHHHHH
DHR16	MNDKAKEAEELLRKALEKAEKENDETAIRCVELLKEALERAKKNNNDKAIEAVELLAKALEKALKENDETAIRCVCLLAEALLRALKNN NDKAIEAVELLAKALEKALKENDETAIRCVCLLAEALLRALKNNNDKAIEEVERLAKELEKALKENDETKIREVCERAEELLRRLKNNN GWLEHHHHHH
DHR17	MSSEDAREKIEQLCREAKEIAERAKQQNSQEEAREAIEKLLRIAKRIAELAKQANQSEVAREAIECLCRIAKLIAELAKQANSQEVARE AIEALLRIAKLIAELAKQANQSEVAREAIECLCRIAKLIAELAKQANSQEVAREAIEALLRIAKLIAELAKQANQSEVAREAIECLSRI AKLIEELAKQANSQEVKREAQEALDRIQKLIEELQKQANQGWLEHHHHHH
DHR18	MDIEKLCKKAESEAREARSKAEELRQRHPDSQAARDAQKLASQAEEAVKLACELAQEHPNADIAKLCIKAASEAAEAASKAAELAQRHP DSQAARDAIKLASQAAEAVKLACELAQEHPNADIAKLCIKAASEAAEAASKAAELAQRHPDSQAARDAIKLASQAAEAVKLACELAQEH PNADIAKKCIKAASEAAEEASKAAEEAQRHPDSQKARDEIKEASQKAEEVKERCERAQEHPNAWLEHHHHHH
DHR19	MDEIEKVREEAEKLKKKTDDEDVLEVAREAIRAAKEATSDEILKVIKEALKLAKKTTDKDVLEVAREAIRAAEEATDDEILKVIKEALK LAKKTTDKDVLEVAREAIRAAEEATDEEILKEIKEALKKAKETTDTEELEKAREQIRKAEESTDGWLEHHHHHH
DHR20	MSDIEEIRQLAEELRKKSDNEEVRKLAQEAAELAKRSTDSDVLEIVKDALELAKQSTNEEVIKLALKAAVLAAKSTDSDVLEIVKDALE LAKQSTNEEVIKLALKAAVLAAKSTDEEVLEEVKEALRRAKESTDEEEIKEELRKAVEEAESTDGWLEHHHHHH
DHR21	MSEKEKVEELAQRIREQLPDTELAREAQELADEARKSDDSEALKVVYLALRIVQQLPDTELAREALELAKEAVKSTDSEALKVVYLALR IVQQLPDTELAREALELAKEAVKSTDQEALKSVYEALQRVQDKPNTEEARESLERAKEDVKSTDGWLEHHHHHH
DHR22	MDDAEELRERARDLLRKNGSSEEEIKKVDEELEKIVRKADSDDAVKLAVKAAALLAENGSSAEEIVKVLEELLKIVEKADSDDAVKLAV KAAALLAENGSSAEEIVKVLEELLKIVEKADSEEEVKDAVREAAELAERGSSAEEIRKQLKDRLRKVEESDSGWLEHHHHHH
DHR23	MSDSEKLAKRVLKELKRRGTSDEELERMKRELEKIIKSATSSDAMRLALRVVLELVRRGTSSEILEKMMRMLIKIIQSATSSDAMRLAL RVVLELVRRGTSSEILEKMMRMLIKIIQSATSDDQMREALRQVLEEVRKGTSSEQLERSMRKLIKEIKKRTSGWLEHHHHHH
DHR24	MSEAEELARRAAKEAKELCKRSTDEELCKELKKLAELLKELAERYPDSEAAKLALKAALEAIELCKQSTDEELCEELVKLAQKLIELAK RYPDSEAAKLALKAALEAIELCKQSTDEELCEELVKLAQKLIELAKRYPDSEEAKRALKEAKELIEQCKESTDEDECRELVKRAEELIR EAKENPDGWLEHHHHHH
DHR25	MDERDKVRELIDRVEKELKREGTSEELIEEIRKVLKKAKEAADSDDDEAIKVAKEIVRVILELVREGTSSELIEEILKVLSLAAEAAKS TDDEAIKVAKEIVRVILELVREGTSSELIEEILKVLSLAAEAAKSTDEEAIKKAKEIVRRILELTREGTSEEEIREELKELRKKAQKAK SPEGWLEHHHHHH
DHR26	MDECERLRQEVEKAEKELEKLAKQSTDEEVRQIAREVAKQLRRLAEEACRSNSDECLRLASEVVKAVQELVKLAEQATDEEVIRVALEV ARELIRLAQEACRSNDDECLRLASEVVKAVQELVKLAEQATDEEVIRVALEVARELIRLAQEACRSNDEECLREASEVVKEVQELVKEA

	EKSTDEEEIRELLQRAEERIREAQERCREGDGWLEHHHHHH
DHR27	MTRQKEQLDEVLEEIQRLAEEARKLMTDEEEAKKIQEEAERAKEMLRRAVEKVTDNEVIEKLLEVVKEIIRLAEEAMKKMTDEEEAAKI AKEALEAIKMLARAVEEVTDNEVIEKLLEVVKEIIRLAEEAMKKMTDEEEAAKIAKEALEAIKMLARAVEEVTDKERIEQLLREVKEEI RRAEEESRKETDDEEAAKRAREALRRIRERAREVEEDKSGWLEHHHHHH
DHR28	MDEEVQRIREEVRRAIEEVRESLERNDSEEAEELAREALERVAEEVKESIKERPDRDLAIEAIRALVRLAIEIVRLALEQNDSELAREV AEEALRAVAEVVKEAIRQRGDRDLAIEAIRALVRLAIEIVRLALEQNDSELAREVAEEALRAVAEVVKEAIRQRGDRELAKEAIRALRR LAEEIRRLAEEQNDDELAREVEELAREAIEEVRKELERQRPGRGWLEHHHHHH
DHR29	MSEVEESAQEVEKRAQEVREEAERRGTSQEVLDEIKRVVDEARQLAQRAKESDDSEVAESALQVVREALKVVLSALERGTSEEVLKEIL RVVSEAIKLALEAIKSSDSEVAESALQVVREALKVVLSALERGTSEEVLKEILRVVSEAIKLALEAIKSSDSETARRALEKVRESLKEV LEQLERGTSEEELRESLREVSENIRKALEEIKSPDGWLEHHHHHH
DHR30	MSTVKELLDRARELMRELAERASEQGSDEEEARKLLEDLEQLVQEIRRELEETGTSSEVIRLIAKAIMLMAELALRAAEQGSDAEEAMK LLKDLLRLVLEILRELRETGTDSEVIRLIAKAIMLMAELALRAAEQGSDAEEAMKLLKDLLRLVLEILRELRETGTDKEEIRKVAEEIM RRAKTALDEARQGSDAEEAMKRLKEQLRRILERLREEREKGTDGWLEHHHHHH
DHR31	MDSYTERARKAVKRYVKEEGGSEEEAEREAEKVREEIRKKASDSYLIQAAAAVVAYVIEEGGSPEEAVKIAEEVVRRIKEKADDSYLIQ AAAAVVAYVIEEGGSPEEAVKIAEEVVRRIKEKADDRELIRRAAERVAEVIERGGSPEEAVKEAEKEVKKQKEESDGWLEHHHHHH
DHR32	MSIQEKAKQSVIRKVKEEGGSEEEARERAKEVEERLKKEADDSTLVRAAAAVVLYVLEKGGSTEEAVQRAREVIERLKKEASDSTLVRA AAAVVLYVLEKGGSTEEAVQRAREVIERLKKEASDEELIREAAKEVLKVLEEGGSVEEAVERARERIEELQKRSDDGWLEHHHHHH
DHR33	MSETEEVKKLVEEKVKKEGGSPEEAKETAKEVTEELKEESQDSTLLKVAALVASAVLKEGGSPEEAAETAKEVVKELRKSASDSTLLKV AALVASAVLKEGGSPEEAAETAKEVVKELRKSASDEELLKEAARQAEESLRQGKSPEEAAEEAKKEVKKLKEKSQDGWLEHHHHHH
DHR34	MSETEEVKKLCEEKVKKEGGSPEEAKETAKEVTEELKEESQDSTLLKVAALCASAVLKEGGSCEEAAETAKEVVKELRKSASDSTLLKV AALCASAVLKEGGSCEEAAETAKEVVKELRKSASDEELLKEAARQAEESLRQGKSCEEAAEEAKKEVKKLKEKSQDGWLEHHHHHH
DHR35	MSEEDEVAKQASRYAKEQGGDPEKSREEAEKALEEVKKQATSSEALQVALEAARYASEEGEDPAEALKEAARALEEVRRSATSSEALQV ALEAARYASEEGEDPAEALKEAARALEEVRRSATSEEDLKEALDRAREASERGQNPAESLKEAAEELKKKKEKSSDGWLEHHHHHH
DHR36	MSDLEKALKRFVKEEKKKGRNPEEAKKEAKKLKKKLKKSAGSSDLLTALAKFVLEEVRKGRNPEEAVKEAIKLAEKLKRSAGSSDLLTA LAKFVLEEVRKGRNPEEAVKEAIKLAEKLKRSAGSSEQLEKLATKVLEEVKKGRNPKRAVEEAIKQAKEDRKRSNSGWLEHHHHHH
DHR37	MSSTERAAQSVKKYLQQQGKDPDQAQKKAQEVKENIEKEANSSSVIRAAAAVVFYLLEQGYDPDQALKKAQEVARNIENEANSSSVIRA AAAVVFYLLEQGYDPDQALKKAQEVARNIENEANSDDVIKEAAKVVYKRLEEGQDPDKALEEARKRAQKTEKKTTSGWLEHHHHHH
DHR38	MSSTERAAQSCKKYLQQQGKDPDQAQKKAQEVKENIEKEANSSSVIRAAAACVFYLLEQGYDCDQALKKAQEVARNIENEANSSSVIRA AAACVFYLLEQGYDCDQALKKAQEVARNIENEANSDDVIKEAAKVVYKRLEEGQDCDKALEEARKRAQKTEKKTTSGWLEHHHHHH
DHR39	MSDLQEVADRIVEQLKREGRSPEEARKEARRLIEEIKQSAGGDSELIEVAVRIVKELEEQGRSPSEAAKEAVELIERIRRAAGGDSELI EVAVRIVKELEEQGRSPSEAAKEAVELIERIRRAAGGDSDRIKKAVELVRELEERGRSPSEAARRAVEEIQRSVEEDGGNGWLEHHHHH H
DHR40	MSESDEVAKRISKEAKKEGRSEEEVKELVERFREAIEKLKEQGDSEAIRVAVEIADEALREGLSPEEVVELVERFVQAIQKLQENGESE AIRVAVEIADEALREGLSPEEVVELVERFVQAIQKLQENGEEDEIQKAVETAQEQLEEGRSPKEVVETVEEQVKEVEEKQQKGEGWLEH HHHHH
DHR41	MSDIEKAKRIADRAIDVVRKAAEKEGGSPEKIREALQQAKRCAEKLIRLVKEAQESNSSDVREAARVALEAVRVVVRAAEEKGGSPEEV VEAVCRAVRCAEKLIRLVKRAEESNSSDVREAARVALEAVRVVVRAAEEKGGSPEEVVEAVCRAVRCAEKLIRLVKRAEESNSENVRES ARRALEKVLKTVQQAEEEGKSPEEVVEQVCRSVRKAEEQIRETQERERSTSGWLEHHHHHH
DHR42	MSDAEEVKKQAEEIANRAYKTAQKQGESDSRAKKAEKLVRKAAEKLARLIERAQKEGDSDALEVARQALEIARRAFETAKKQGHSATEA AKAFVDVVEAAISLAELIISAKRQGDSDALEVARQALEIARRAFETAKKQGHSATEAAKAFVDVVEAAISLAELIISAKRQGDQKALEI ARKALQKAKENFEEAQKRGESATQAAKRFVDTVEKEIKKAQEQIKRERKGDGWLEHHHHHH
DHR43	MSKEEELIEKARRVAKEAIEEAKRQGKDPSEAKKAAEKLIKAVEEAVKEAKRLKEEGNSELAELISEAIQVAVEAVEEAVRQGKDPFKA AEAAAELIRAVVEAVKEAERLKREGNSELAELISEAIQVAVEAVEEAVRQGKDPFKAAEAAAELIRAVVEAVKEAERLKREGNSELAKK INDTIREAVREVQQAVEDGKDPFEAAREAAEKIRESVERVREEEEKKRRGNGWLEHHHHHH
DHR44	MSNEQEKKDLKKAEEAAKSPDPELIREAIERAEESGSNKAKEIILRAAEEAAKSPDPELIRLAIEAAERSGSNKAKEIILRAAEEAAKS PDPELIRLAIEAAERSGSEKAKEIIKRAAEEAQKSPDPELQKLAKEARERLGGWLEHHHHHH
DHR45	MSSEEEELEKDAREASESGADPEWLREIVDLARESGDSEVIELAKRALEAAKSGADPEWLLRIVRQAEESGSSEVIELAKRALEAAKSG ADPEWLLRIVRQAEESGSEEVIELAKRALEEAKKGKDPKELLEEVRKREESGGWLEHHHHHH
DHR46	MSTKEEKERIERIEKEVRSPDPENIREAVRKAEELLRENPSTEAEELLRRAIEAAVRAPDPEAIREAVRAAEELLRENPSTEAEELLRR AIEAAVRAPDPEAIREAVRAAEELLRENPSEEAKELLRRAIESAKKAPDPEAQREAKRAEEELRKEDPGWLEHHHHHH
DHR47	MSTKEEKERIERIEKEVRSPDCENIREAVRKAEELLRENPSTEAEELLRRAIEAAVRCPDCEAIREAVRAAEELLRENPSTEAEELLRR AIEAAVRCPDCEAIREAVRAAEELLRENPSEEAKELLRRAIESAKKCPDPEAQREAKRAEEELRKEDPGWLEHHHHHH
DHR48	MNSREEEEAKRIVKEAKKSGFDPEEVEKALREVIRVAEETGNSEALKEALKIVEEAAKSGYDPAEVAKALAEVIRVAEETGNSEALKEA LKIVEEAAKSGYDPAEVAKALAEVIRVAEETGNPEELKEALKRVLEAAKRGEDPAQVAKELAEEIRRNQEEGGWLEHHHHHH
DHR49	MDSEEEQERIRRILKEARKSGTEESLRQAIEDVAQLAKKSQDSEVLEEAIRVILRIAKESGSEEALRQAIRAVAEIAKEAQDSEVLEEA IRVILRIAKESGSEEALRQAIRAVAEIAKEAQDPRVLEEAIRVIRQIAEESGSEEARRQAERAEEEIRRRAQGWLEHHHHHH
DHR50	MDPEEVRREVERATEEYRKNPGSDEAREQLKEAVERAEEAARSPDPEAVQVAVEAATQIYENTPGSEEAKKALEIAVRAAENAARLPDP EAVQVAVEAATQIYENTPGSEEAKKALEIAVRAAENAARLPDPEAVRVAEEAADQIRKNTPGSELAKRADEIKKRARELLERLPGWLEH HHHHH
DHR51	MQSEDRKEKIRELERKARENTGSDEARQAVKEIARIAKEALEEGNADTAKEAIQRLEDLARDYSGSDVASLAVKAIAKIAETALRNGYA DTAKEAIQRLEDLARDYSGSDVASLAVKAIAKIAETALRNGYKETAEEAIKRLRELAEDYKGSEVAKLAEEAIERIEKVSRERGGWLEH HHHHH
DHR52	MQCEDRKEKIRELERKARENTGSDEARQAVKEIARIAKEALEEGCCDTAKEAIQRLEDLARDYSGSDVASLAVKAIAKIAETALRNGCC DTAKEAIQRLEDLARDYSGSDVASLAVKAIAKIAETALRNGCKETAEEAIKRLRELAEDYKGSEVAKLAEEAIERIEKVSRERGGWLEH HHHHH
DHR53	MSNDEKEKLKELLKRAEELAKSPDPEDLKEAVRLAEEVVRERPGSNLAKKALEIILRAAEELAKLPDPEALKEAVKAAEKVVREQPGSN LAKKALEIILRAAEELAKLPDPEALKEAVKAAEKVVREQPGSELAKKALEIIERAAEELKKSPDPEAQKEAKKAEQKVREERPGGWLEH HHHHH
DHR54	MTTEDERRELEKVARKAIEAAREGNTDEVREQLQRALEIARESGTTEAVKLALEVVARVAIEAARRGNTDAVREALEVALEIARESGTT EAVKLALEVVARVAIEAARRGNTDAVREALEVALEIARESGTEEAVRLALEVVKRVSDEAKKQGNEDAVKEAEEVRKKIEEESGGWLEH HHHHH

DHR55	MSSVAEEIEKRAKKISKELKKEGKNPEWIEELQRAADKLVEVARRATSSDALEIAKRAVKIAEELAKQGSNPKWIAELLKAAAKLVEVA ARATSSDALEIAKRAVKIAEELAKQGSNPKWIAELLKAAAKLVEVAARATSPKALKQAKEAVKEAEELAKKGRNPKEIAEELKKRAKEV EKLARSTGWLEHHHHHH
DHR56	MSSVAEEIEKRCKKISKELKKEGKNPEWIEELQRACDKLVEVARRATSSDALEIAKRCVKIAEELAKQGSNPKWIAELLKACAKLVEVA ARATSSDALEIAKRCVKIAEELAKQGSNPKWIAELLKACAKLVEVAARATSPKALKQAKECVKEAEELAKKGRNPKEIAEELKKCAKEV EKLARSTGWLEHHHHHH
DHR57	MSTEELKKVLERVRELSERAKESTDPEEALKIAKEVIELALKAVKEDPSTDALRAVLEAVRLASEVAKRVTDPDKALKIAKLVIELALE AVKEDPSTDALRAVLEAVRLASEVAKRVTDPDKALKIAKLVIELALEAVKEDPSEEAKRAVEEAKRLAEEVSKRVTDPELSEKIRQLVK ELEEEAQKEDPGWLEHHHHHH
DHR58	MSTEELKKVLERVRELCERAKESTDPEEALKIAKEVIELALKAVKEDPSTDALRAVLEAVRCACEVAKRVTDPDKALKIAKLVIELALE AVKEDPSTDALRAVLEAVRCACEVAKRVTDPDKALKIAKLVIELALEAVKEDPSEEAKRAVEEAKRCAEEVSKRVTDPELSEKIRQLVK ELEEEAQKEDPGWLEHHHHHH
DHR59	MKTEVEKKAKEVIKEAKELAKELDSEEAKKVVERIKEAAEAAKRAAEQGKTEVAKLALKVLEEAIELAKENRSEEALKVVLEIARAALA AAQAAEEGKTEVAKLALKVLEEAIELAKENRSEEALKVVLEIARAALAAAQAAEEGKSDEARDALRRLEEAIEEAKENRSKESLEKVRE EAKEAEQQAEDAREGGWLEHHHHHH
DHR60	MTDIKKKAEEIIKEAKKQGSEDAIRLAQEAKKQGTDILVRAAEIVVRAQEQGSEDAIRLAKEASREGTDILVRAAEIVVRAQEQGSEDA IRLAKEASREGTPTLVKAAEKVVRAQQKGSQDTIEKAKEESREGGWLEHHHHHH
DHR61	MTDIKKKAEEIIKEAKKQGSEDAIRLAQECKKQGTDICVRAAEIVVRAQEQGSEDAIRLAKECSREGTDICVRAAEIVVRAQEQGSEDA IRLAKECSREGTPTCVKAAEKVVRAQQKGSQDTIEKAKEESREGGWLEHHHHHH
DHR62	MDNDEKRKRAEKALQRAQEAEKKGDVEEAVRAAQEAVRAAKESGDNDVLRKVAEQALRIAKEAEKQGNVEVAVKAARVAVEAAKQAGDN DVLRKVAEQALRIAKEAEKQGNVEVAVKAARVAVEAAKQAGDQDVLRKVSEQAERISKEAKKQGNSEVSEEARKVADEAKKQTGGWLEH HHHHH
DHR63	MDPDEDRERLKEELKKIREALREAKEKPDPEEIKRALREVLEAIRRILKLAERAGDPDLAREALKEINKVIREALEIAKRVPDPEVIKE ALRVVLEAIRAILKLAEQAGDPDLAREALKEINKVIREALEIAKRVPDPEVIKEALRVVLEAIRAILKLAEQAGDPDLAREALEEIDKV IDEAQEISERVPDEEVQREAQEVIKEADRARKKLSEQSGGWLEHHHHHH
DHR64	MDPEDELKRVEKLVKEAEELLRQAKEKGSEEDLEKALRTAEEAAREAKKVLEQAEKEGDPEVALRAVELVVRVAELLLRIAKESGSEEA LERALRVAEEAARLAKRVLELAEKQGDPEVALRAVELVVRVAELLLRIAKESGSEEALERALRVAEEAARLAKRVLELAEKQGDPEVAR RAVELVKRVAELLERIARESGSEEAKERAERVREEARELQERVKELREREGGWLEHHHHHH
DHR65	MDPEDELKRVEKLVKEAEELLRQCKEKGSEECLEKALRTAEEAAREAKKVLEQAEKEGDPEVALRAVELVVRVAELLLRICKESGSEEC LERALRVAEEAARLAKRVLELAEKQGDPEVALRAVELVVRVAELLLRICKESGSEECLERALRVAEEAARLAKRVLELAEKQGDPEVAR RAVELVKRVAELLERICRESGSEECKERAERVREEARELQERVKELREREGGWLEHHHHHH
DHR66	MTSDDDKVREAEERVREAIERIQRALKKRDTPDARKALEAAKKLLKVVEKAKKRGTSDAIKVAEAAARVAEAIARILEALNERDTPDAR KALRAAIKLAEVVYKAAESGTSDAIKVAEAAARVAEAIARILEALNERDTPDARKALRAAIKLAEVVYKAAESGTTEALKVAEKAARVA EKIARILEKLNERDTPEARKKLRQAIKEAEKVYKESEQGGWLEHHHHHH
DHR67	MTSEIDKLIKKLRQTAKEVKREAEERKRRSTDPTVREVIERLAQLALDVAEEAARLIKKATTSEVAKLVWKLARTAIEVIREAIERAER STDPEVIRVILELARLAAEVAKEAARLIVKATTSEVAKLVWKLARTAIEVIREAIERAERSTDPEVIRVILELARLAAEVAKEAARLIV KATTEEVAKKVWKEAYRAIEEIRKAIEKAERSTDPNEIKKILEEARKKAEEAIERAKEIVKSTGWLEHHHHHH
DHR68	MTPRERLEEAKERVEEIRELIDKARKLQEQGNKEEAEKVLREAREQIREVTRELEEIAKNSDTPELALRAAELLVRLIKLLIEIAKLLQ EQGNKEEAEKVLREATELIKRVTELLEKIAKNSDTPELALRAAELLVRLIKLLIEIAKLLQEQGNKEEAEKVLREATELIKRVTELLEK IAKNSDTPELAKRAAELLKRLIELLKEIAKLLEEEGNEDEAEKVKEEAKELEERVRELEERIRKNSDGWLEHHHHHH
DHR69	MNPQEDLERAEKVVRSVEEVLQRAKEAQREGDKEKVERLIKEAENQIRKARELLERVVRQNPDDPEVLLRVAELIVRLVEVVLELAKLA EKNGDKEQVERLIQTAEELIREARELLERVSREIPDNPEVLLRVAELIVRLVEVVLELAKLAEKNGDKEQVERLIQTAEELIREARELL ERVSREIPDNPESLKRVAELIKRLVKVVDELSKLAERNGDRDQVERLRQLAEELRREAEELEERVRRERPDGWLEHHHHHH
DHR70	MSTEEKIEEARQSIKEAERSLREGNPEKAREDVRRALELVRELEKLARKTGSTEVLIEAARLAIEVARVALKVGSPETAREAVRTALEL VQELERQARKTGSTEVLIEAARLAIEVARVALKVGSPETAREAVRTALELVQELERQARKTGSDEVLKRAAELAKEVARVAKEVGSPET ARQARETAERLREELRRNREKKGGWLEHHHHHH
DHR71	MDPEEILERAKESLERAREASERGDEEEFRKAAEKALELAKRLVEQAKKEGDPELVLEAAKVALRVAELAAKNGDKEVFKKAAESALEV AKRLVEVASKEGDPELVLEAAKVALRVAELAAKNGDKEVFKKAAESALEVAKRLVEVASKEGDPELVEEAAKVAEEVRKLAKKQGDEEV YEKARETAREVKEELKRVREEKGGWLEHHHHHH
DHR72	MDSTKEKARQLAEEAKETAEKVGDPELIKLAEQASQEGDSEKAKAILLAAEAARVAKEVGDPELIKLALEAARRGDSEKAKAILLAAEA ARVAKEVGDPELIKLALEAARRGDSEKARAILEAAERAREAKERGDPEQIKKARELAKRGGWLEHHHHHH
DHR73	MDAEEEAKEAIKRAQEAIELARKGNPEEARKVAEEARERAERVREEAEKRGDAEVLALVAIALALVAIALAEVGNPEEAREVAERAKEI AERVRELAEKRGDAEVLALVAIALALVAIALAEVGNPEEAREVAERAKEIAERVRELAEKRGDARVLKLVAKALELVAEALKKVGNPEE AREVEERAREIKERVRRLLEEKGGWLEHHHHHH
DHR74	MDSEADRIIKKLQKEIKEVEQEARDSNDDEERELLKRLAEALKRAAEAVKRAQESGDSEAIRIIKKLVKEITEVVREARKSTDKEEIEL LIRLAEALARAAEAVADAAKSGDSEAIRIIKKLVKEITEVVREARKSTDKEEIELLIRLAEALARAAEAVADAAKSGDQEAIKRIKKLV KKIIEVVRKARKSTNKKEIEKLIRKAEKLARKAEQIAEDAKRGGWLEHHHHHH
DHR75	MDSEKEKATELAERAQDVASRVEEEARREGSRELIEIARELRERAEEASQEGDSEKAKAILLAAKAVLVAVEVYERAKRQGSDELREIA RELAKEALRAAQEGDSEKAKAILLAAKAVLVAVEVYERAKRQGSDELREIARELAKEALRAAQEGDSEKARAILEAAREVLRAVEQYER AKRRGDDDERERAREEAREALERAREGGWLEHHHHHH
DHR76	MNPELEEWIRRAKEVAKEVEKVAQRAEEEGNPDLRDSAKELRRAVEEAIEEAKKQGNPELVEWVARAAKVAAEVIKVAIQAEKEGNRDL FRAALELVRAVIEAIEEAVKQGNPELVEWVARAAKVAAEVIKVAIQAEKEGNRDLFRAALELVRAVIEAIEEAVKQGNPELVERVARLA KKAAELIKRAIRAEKEGNRDERREALERVREVIERIEELVRQGGWLEHHHHHH
DHR77	MNSDEEEAREWAERAEEAAKEALEQAKREGDEDARRVAEELEKQAEEARRKKDSEEAEAVYWAARAVLAALEALEQAKREGDEDARRVA EELLRQAEEAARKKNSEEAEAVYWAARAVLAALEALEQAKREGDEDARRVAEELLRQAEEAARKKNPEEARAVYEAARDVLEALQRLEE AKRRGDEEERREAEERLRQAEERARKKGWLEHHHHHH
DHR78	MNSDEEEAREWAERAEEAAKEALEQAKREGDEDARRCAEELEKQAEEARRKKDSEEAEAVYWAARAVLAALEALEQAKREGDEDARRCA EELLRQACEAARKKNSEEAEAVYWAARAVLAALEALEQAKREGDEDARRCAEELLRQACEAARKKNPEEARAVYEAARDVLEALQRLEE AKRRGDEEERREAEERLRQACERARKKGWLEHHHHHH
DHR79	MSSDEEEARELIERAKEAAERAQEAAERTGDPRVRELARELKRLAQEAAEEVKRDPSSSDVNEALKLIVEAIEAAVRALEAAERTGDPE VRELARELVRLAVEAAEEVQRNPSSSDVNEALKLIVEAIEAAVRALEAAERTGDPEVRELARELVRLAVEAAEEVQRNPSSEEVNEALK

	KIVKAIQEAVESLREAEESGDPEKREKARERVREAVERAEEVQRDPSGWLEHHHHHH
DHR80	MNSEELERESEEAERRLQEARKRSEEARERGDLKELAEALIEEARAVQELARVASERGNSEEAERASEKAQRVLEEARKVSEEAREQGD DEVLALALIAIALAVLALAEVASSRGNSEEAERASEKAQRVLEEARKVSEEAREQGDDEVLALALIAIALAVLALAEVASSRGNKEEAE RAYEDARRVEEEARKVKESAEEQGDSEVKRLAEEAEQLAREARRHVQETRGGWLEHHHHHH
DHR81	MNSEELERESEEAERRLQEARKRSEEARERGDLKELAEALIEEARAVQELARVACERGNSEEAERASEKAQRVLEEARKVSEEAREQGD DEVLALALIAIALAVLALAEVACCRGNSEEAERASEKAQRVLEEARKVSEEAREQGDDEVLALALIAIALAVLALAEVACCRGNKEEAE RAYEDARRVEEEARKVKESAEEQGDSEVKRLAEEAEQLAREARRHVQECRGGWLEHHHHHH
DHR82	MNDEEVQEAVERAEELREEAEELIKKARKTGDPELLRKALEALEEAVRAVEEAIKRNPDNDEAVETAVRLARELKKVAEELQERAKKTG DPELLKLALRALEVAVRAVELAIKSNPDNDEAVETAVRLARELKKVAEELQERAKKTGDPELLKLALRALEVAVRAVELAIKSNPDNEE AVETAKRLAEELRKVAELLEERAKETGDPELQELAKRAKEVADRARELAKKSNPNGWLEHHHHHH
DHR83	MNDEEVQEACERAEELREEAEELIKKARKTGDPELLRKALEALEEAVRAVEEAIKRNPDNDECVETACRLARELKKVAEELQERAKKTG DPELLKLALRALEVAVRAVELAIKSNPDNDECVETACRLARELKKVAEELQERAKKTGDPELLKLALRALEVAVRAVELAIKSNPDNEE CVETAKRLAEELRKVAELLEERAKETGDPELQELAKRAKEVADRARELAKKSNPNGWLEHHHHHH

Supplementary References

1. Leaver-Fay, A. *et al.* ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.* **487,** 545–574 (2011).
2. Rohl, C. A., Strauss, C. E. M., Misura, K. M. S. & Baker, D. in *Methods in Enzymology* (ed. Johnson, L. B. and M. L.) **383,** 66–93 (Academic Press, 2004).
3. Wintjens, R. T., Rooman, M. J. & Wodak, S. J. Automatic Classification and Analysis of αα-Turn Motifs in Proteins. *J. Mol. Biol.* **255,** 235–253 (1996).
4. Koga, N. *et al.* Principles for designing ideal protein structures. *Nature* **491,** 222–227 (2012).
5. Kuhlman, B. & Baker, D. Native protein sequences are close to optimal for their structures. *Proc. Natl. Acad. Sci.* **97,** 10383–10388 (2000).
6. Sheffler, W. & Baker, D. RosettaHoles: Rapid assessment of protein core packing for structure prediction, refinement, design, and validation. *Protein Sci.* **18,** 229–239 (2009).
7. Parmeggiani, F. *et al.* A General Computational Approach for Repeat Protein Design. *J. Mol. Biol.* **427,** 563–575 (2015).
8. Kuhlman, B. *et al.* Design of a Novel Globular Protein Fold with Atomic-Level Accuracy. *Science* **302,** 1364–1368 (2003).
9. Jones, D. T. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292,** 195–202 (1999).
10. Bradley, P., Misura, K. M. S. & Baker, D. Toward High-Resolution *de Novo* Structure Prediction for Small Proteins. *Science* **309,** 1868–1871 (2005).
11. Das, R. *et al.* Structure prediction for CASP7 targets using extensive all-atom refinement with Rosetta@home. *Proteins Struct. Funct. Bioinforma.* **69,** 118–128 (2007).
12. Huang, P.-S. *et al.* RosettaRemodel: A Generalized Framework for Flexible Backbone Protein Design. *PLoS ONE* **6,** e24109 (2011).

13. Kajava, A. V. Tandem repeats in proteins: From sequence to structure. *J. Struct. Biol.* **179,** 279–288 (2012).
14. Di Domenico, T. *et al.* RepeatsDB: a database of tandem repeat protein structures. *Nucleic Acids Res.* **42,** D352–D357 (2014).
15. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22,** 1658–1659 (2006).
16. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28,** 3150–3152 (2012).
17. Zhang, Y. & Skolnick, J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* **33,** 2302–2309 (2005).
18. Walsh, I. *et al.* RAPHAEL: recognition, periodicity and insertion assignment of solenoid protein structures. *Bioinformatics* **28,** 3257–3264 (2012).
19. Punta, M. *et al.* The Pfam protein families database. *Nucleic Acids Res.* **40,** D290–D301 (2012).
20. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22,** 2577–2637 (1983).
21. Kobe, B. & Kajava, A. V. When protein folding is simplified to protein coiling: the continuum of solenoid protein structures. *Trends Biochem. Sci.* **25,** 509–515 (2000).
22. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–3402 (1997).
23. Remmert, M., Biegert, A., Hauser, A. & Söding, J. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat. Methods* **9,** 173–175 (2012).

24. Weiss, M. S. & Hilgenfeld, R. On the use of the merging R factor as a quality indicator for X-ray data. *J. Appl. Crystallogr.* **30,** 203–205 (1997).
25. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D* **66,** 12–21 (2010).
26. Petoukhov, M. V. *et al.* New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* **45,** 342–350 (2012).
27. Classen, S. *et al.* Implementation and performance of SIBYLS: a dual endstation small-angle X-ray scattering and macromolecular crystallography beamline at the Advanced Light Source. *J. Appl. Crystallogr.* **46,** 1–13 (2013).
28. Rambo, R. P. & Tainer, J. A. Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. *Biopolymers* **95,** 559–571 (2011).
29. Rambo, R. P. & Tainer, J. A. Accurate assessment of mass, models and resolution by small-angle scattering. *Nature* **496,** 477–481 (2013).